சொற்பொருண்மை மயக்கநீக்கம்: கணினிசார்அணுகுமுறை

ச. இராசேந்திரன்
அமிர்தா பல்கலைகழகம்
கோயம்புத்தூர்

முன்னுரை

மக்கள் பொருண்மை மயக்கம் காட்டுகின்றார்கள் என்று கூறுவது அவர்கள் எவ்வாறு மொழியைப் பயன்படுத்துகின்றனர் என்பதைக் குறிக்கும் என்றாலும், பொருண்மை மயக்கம் என்பது மொழி வெளிப்பாடுகளின் பண்பாகும். ஒரு சொல்லோ தொடரோ வாக்கியமோ ஒன்றிற்கும் மேற்பட்ட பொருண்மைகளைக் கொண்டிருந்தால் அவை பொருண்மை மயக்கம் கொண்டுள்ளன என்று பொருள்படும். வெளிப்படையாக இந்த வரையறை விளக்கம் எந்தப் பொருள்கள் அல்லது ஒரு வெளிப்பாடு ஒன்றைக் கொண்டிருக்கும் என்பது என்ன என்று கூறவில்லை. ஒரு குறிப்பிட்ட மொழிக்கு இந்தத் தகவல் இலக்கணத்தால் தரப்படும்; இது முறையாக வடிவங்களைப் பொருண்மைகளுடன் பொருத்தும் மற்றும் பொருண்மை மயக்கம் உள்ள வடிவங்களை ஒன்றிற்கும் மேற்பட்ட பொருண்மைகளுடன் பொருத்தும்.

சொற்பொருண்மை மயக்கம் தொடரிலும் வாக்கியத்திலும் பொருண்மை மயக்கத்தை உருவாக்கும். எடுத்துக்காட்டாக, *கால் பகுதி.* இருப்பினும் தொடரும் வாக்கியமும் சொற்பொருண்மை மயக்கம் உள்ள சொல் இல்லாவிடினும் பொருள் மயக்கம் தரலாம். எடுத்துக்காட்டாகப் பின்வரும் தொடர்களை எடுத்துக்கொள்ளவும்.

(1) வெள்ளை மருந்துக் குப்பி
(2) பழைய மாணவர் விடுதி

முதல் வாக்கியம் 'மருந்துள்ள வெள்ளை நிறக் குப்பி' என்றும் 'வெள்ளை மருந்துள்ள குப்பி' என்றும் பொருள் மயக்கம் காட்டும். இதுபோன்று இரண்டாவது வாக்கியமும் 'பழைய மாணவர்கள் வசிக்கும் விடுதி' என்றும் 'மாணவர்கள் வசிக்கும் பழைய விடுதி' என்றும் பொருள் மயக்கம் தரும். இத்தகைய பொருண்மை மயக்கம் அமைப்புப் பொருண்மை மயக்கம் எனப்படும்.

இருவகையான பொருண்மை மயக்கங்கள் முக்கியமாகப் பேசப்படுகின்றன. அவையாவன சொற்பொருண்மை மயக்கம் மற்றும் அமைப்புப் பொருண்மை மயக்கம். இக்கட்டுரையில் சொற்பொருண்மை மயக்கம் குறித்தே ஆய இருக்கின்றோம்.

1. சொற்பொருண்மை மயக்கம்

சொற்பொருண்மை மயக்கம் மிகப் பொதுவானதாகும். எடுத்துக்காட்டுகள் *படி, குடி, மடம்* என்ற பெயர்களையும் *பிடி, அறை, முடி, சுடு* என்ற வினைகளையும் *நீண்ட, விரிந்த, குறைந்த, வெளுத்த, கறுத்த* என்ற பெயரடைகளையும் உள்ளடக்கும். பொருண்மை மயக்கத்திற்கு பல சோதனைகள் உள்ளன. ஒரு சோதனை *கடினமான* என்பதற்கு *மெதுவான, எளிதான* என்ற இரண்டும் எதிர்சொற்கள் ஆகும். பின்வரும் வாக்கியத்தைக் கருத்தில் கொள்ளவும். *கடினமான மிட்டாயைக் கடிக்க முடியாது, கடினமான சொல்லுக்குப் பொருள் கூற இயலாது*. முன்னர் தந்த பொருள் மயக்கத்திற்கான எடுத்துக்காட்டு ஒரு சொல்லுக்கு ஒன்றுக்கும் மேற்பட்ட பொருண்மை காரணமாக விளைந்தது. இருப்பினும் எப்போது நமக்கு ஒரு சொல் மட்டும் இருக்கும் என்பதில் தெளிவில்லை. *படி* என்ற வினையும் *படி* என்ற பெயரும் ஒரே உச்சரிப்பைக் கொண்டிருந்தாலும் இரண்டு சொற்கள் ஆகும். இவை ஒப்புருச் சொல்லுக்கு எடுத்துக்காட்டாகும். *கடி* என்ற வினையும் *கடி* என்ற பெயரும் ஒப்புருச் சொற்களா? *முதல் மாணவனில்* வரும் *முதல்* என்பதும் *ஐந்து முதல்* என்பதில் வரும் *முதல்* என்பதும் ஒன்றா? என்ற கேள்விகள் எப்பொழுதும் எழுப்பப்படும். ஒன்று பொருள் மயக்கம் காட்டுவது மற்றொன்று ஒப்புருச் சொற்கள் என்று கூறுவதில் ஒருமித்த கருத்து இல்லை. ஒரு வேளை இந்த வேறுபாடு தன்னிச்சையானதாகும்.

சில வேளைகளில் ஒரு சொல்லின் பொருள் பிற சொல்லிலிருந்து ஆக்கப்பட்டிருக்கும். எடுத்துக்காட்டாக *காண்* என்ற வினைச் சொலின் பொருள் *கண்* என்ற சொல்லின் பொருளிலிருந்து உருவாகி இருக்கக்கூடும். இதே போன்று *எரி, உடை, ஒடி* என்ற சொற்களின் செயப்படுபொருள் குன்றாவினைப் பொருண்மைகள் செயப்படுபொருள் குன்றிய வினைப்பொருண்மைகளிலிருந்து ஆக்கப்பட்டிருக்கலாம். இப்போது இந்த நேர்வுகள் ஒவ்வொன்றிலும் ஆக்கப்பட்ட பொருண்மை அச்சொல்லின் இரண்டாவது பொருண்மையாகத் தகுதி பொறாது; ஆனால் இது ஆக்கப்படாத பொருண்மையின் மீதான சொல்லியல்சார் செயற்பாடாகும். இந்த நிகழ்வு குறிப்பிட்ட சொற்களுக்குச் சிறப்பானது என்றில்லாமல் முறையானது மற்றும் பொதுவானது என்ற அளவில் இந்த வாதம் சாத்தியமானதாகும். சொற்பொருண்மை இயலுக்கு இம்மாதிரியான பொருண்மை நிகழ்வுகளை அடையாளம் காணும் மற்றும் பண்பாக்கம் செய்யும் பொறுப்பு உள்ளது. மேலும் *செய், போடு* என்ற வினைகளும் *மேல், கீழ்* என்ற பின்னுருபுகளும் காட்டும் பொதுவான மற்றும் மிக நெகிழ்வான பொருண்மை நடத்தையை விளக்கும் பொறுப்பும் உள்ளது.

2. சொற்பொருண்மை மயக்கத்தின் வகைகள்

சொற்பொருண்மை மயக்கங்கள் மூன்று அடிப்படை வகைகளில் வேறு படும்: சொல்வகைப்பாட்டுப் பொருண்மை மயக்கம் (parts-of-speech or category ambiguity), ஒப்புருச்சொன்மைசார் (சொற்போலி) பொருண்மை மயக்கம் (ambiguity due to homography),

பல்பொருள் ஒருமொழியம்சார் (பல்பொருண்மை) பொருண்மை மயக்கம் (ambiguity due to polysemy), மாற்றப் பொருண்மை மயக்கம் (ambiguity due to transfer or transfer ambiguity).

2.1 சொல்வகைப்பாட்டுப் பொருண்மை மயக்கம்

சொற்பொருண்மை மயக்கத்தின் நேரடியான வகை சொல்வகைப்பாட்டுப் பொருண்மை மயக்கம் ஆகும்: ஒரு சொல் ஒன்றிற்கும் மேற்பட்ட இலக்கண அல்லது தொடரியல் வகைப்பாட்டிற்கு ஒதுக்கப்படும். தமிழில் இத்தகைய பல எடுத்துகாட்டுகள் உள்ளன. *பச்சை* என்பது பெயராகவும் பெயரடையாகவும் வரும்; *சுடு* என்பது வினையாகவும் பெயரடையாகவும் வரும்; *கடி* என்று வினையாகவும் பெயராகவும் வரும். *மேலே, கீழே* என்ற சொற்கள் பெயராகவும் வினையடைகளாகவும் பின்னுருபுகளாகவும் வரும்.

(3) அவன் மேலே இருக்கிறான். (பெயர்)

(4) அவன் மேலே சென்றான். (வினையடை)

(5) அவன் மேசை மேலே நிற்கின்றான். (பின்னுருபு)

2.2 ஒப்புருச்சொன்மை சார், பல்பொருள் ஒருமொழியம் சார் பொருண்மை மயக்கம்

ஒரு சொல்லுக்கு இரண்டோ அதற்கு மேலோ பொருண்மைகள் இருந்தால் இரண்டாவது வகைச் சொற்பொருண்மை மயக்கம் ஏற்படும். இரண்டோ அதற்கு மேற்பட்ட சொற்கள் முற்றிலும் வேறுபட்ட பொருண்மையைக் கொண்டிருப்பது ஒப்புருச்சொன்மை (homography) ஆகும். எடுத்துகாட்டாகப் *படி* என்பது மாடிப்படியையும் கற்பதையும் குறிப்பிடும்.

(6) அவன் தந்தை அவனிடம் நீ நன்றாகப் படி என்று கூறினார்.

(7) அவன் படி வழியாக மடியில் ஏறினான்.

அட்டை என்பது புத்தக அட்டையையும் அட்டைப் பூச்சியையும் குறிப்பிடும்.

(8) அவன் புத்தகத்தின் அட்டையைக் கிழித்து எறிந்தான்.

(9) அவன் அந்த அட்டையைக் கொன்றான்.

பல்பொருள் ஒருமொழிகள் (polysememes) தொடர்புள்ள ஒரு தொடர்ச்சியான பொருண்மைகளை வெளிப்படுத்தும். சொற்கள் தமது உருவக நீட்சியாலும் மாற்றத்தாலும் புதிய பொருண்மைகளை வெளிப்படுத்தும். எடுத்துக்காட்டாகக் *கிளை* என்பது மரக்கிளையையும் வங்கிக் கிளையையும் குறிப்பிடும்; *நட* எனபது நடக்கும் இயக்கத்தையும் நிறுவனத்தின் செயல்பாட்டையும் குறிப்பிடும்.

(10) தினமும் காலையில் பள்ளிக்கு நடந்து செல்கின்றான்.

(11) அந்த நிறுவனம் நன்றாக நடந்து கொண்டிருக்கின்றது.

ஓடு என்பது மனித இயக்கத்தையும் ஆறு ஒழுகுவதையும் குறிப்பிடும்.

(12) அவன் விரைவாக ஓடுகிறான்.

(13) காவிரி ஆறு தஞ்சாவூர் வழியாக ஓடுகின்றது.

கண் என்பது விலங்கினங்களின் கண்ணையும் தேங்காய் கண்ணையும் குறிப்பிடும்.

(14) அவன் தன் கண்களை மூடினான்.

(15) தேங்காய்க்கு மூன்று கண்கள் உண்டு.

பின்வரும் வாக்கியதில் *கேள்* என்ற சொல் செவிப்புலத்தையும் வினவுதலையும் குறித்து நின்று சொற்பொருண்மை மயக்கம் காட்டும்.

(16) ராதை ராஜா கேட்டதை அவனிடம் கூறினாள்.

இது சொற்பொருண்மைஅடிப்படையில் இரு பொருள்களைத் தந்து சொற்பொருண்மை மயக்கம் காட்டும்.

(17) ராதை ராஜா தன் காதால் கேட்டதை அவனிடம் கூறினாள்.

(18) ராதை ராஜா வினவியதை அவனிடம் கூறினாள்.

சில வேளைகளில் ஒப்புருச் சொல்லிணைகளில் ஒன்றின் பயன்பாடு மற்றதைவிட கூடுதலாக இருக்கக்கூடும். இந்நேர்வில் ஒப்புருச்சொற்களின் பொருண்மை மயக்கத்தை பனுவல் வகை அடிப்படையில் நீக்க இயலும்; எனவே வழக்கமில்லாத பயன்பாடு மொழிபெயர்க்க வேண்டிய பனுவலின் விசயத்திற்குப் பொருந்தினால் ஒழிய அகராதியிலிருந்து விலக்கி வைக்கப்படும்.

இயந்திர மொழிபெயர்ப்பில் ஒப்புருச்சொன்மையையும் (homography) பல்பொருள் ஒருமொழியத்தையும் (polysemy) பெரும்பாலும் ஒன்றுப்போல் கையாள இயலும். ஏனென்றால் இங்கு நோக்கம் எழுத்தப்பட்ட சொல்லுக்குச் சூழல் அடிப்படையில் பொருண்மையைக் கண்டுபிடிப்பது ஆகும். மேற் சொன்ன வழியில் வேறுபட்ட சொல்வகைப்பாட்டைச் சார்ந்த ஒப்புருச் சொற்களின் பொருண்மை மயக்கத்தைத் நீக்க இயலும்; ஆனால் ஒப்புருச் சொற்கள் ஒரே சொல்வகைப்பாட்டைச் சார்ந்திருந்தால் தொடரியல் பகுப்பாய்வு மட்டும் போதுமானதல்ல. பொதுவான ஒரு அணுகுமுறை ‘மனிதவினம்’, ‘பெண்ணினம்’, ‘திரவம்’ போன்ற பொருண்மைப்

பண்புக்கூறுகளைத் தந்து தேர்வுக்கட்டுப்பாடு அடிப்படையில் எந்தப் பண்புக்கூறு தரப்பட்ட தொடரியல் கட்டுமானத்துடன் இணங்கும் என்று குறிப்பிடுவதாகும். எடுத்துக்காட்டாக *குடி* என்ற வினை ஒரு 'விலங்கின' எழுவாயைக் கொண்டிருக்கவேண்டும்.

தொடர்ந்து பயன்படுத்த இயலும் பொருண்மைப் பண்புகூறுகளின் குழுமத்தை உருவாக்குவதுதிலும் இப்பண்புக்கூறுகள் அடிப்படையில் பெயர்கள் மற்றும் வினைகள் இவற்றின் தேர்வுக் கட்டுப்பாடுகளைக் குறிப்பிடுவதிலும் சிக்கல்கள் உள்ளன. இருப்பினும் அவைகள் மொழிபெயர்ப்பு ஒழுங்கு முறைகளில் பெரும்பாலும் வேற்றுமைப் பங்கெடுப்பாளர்கள் (case roles) தகவலுடன் சேர்த்துப் பரவலாகப் பயன்படுத்தப்படுகின்றன. எடுத்துக்காட்டாக *அட்டை* என்ற ஒப்புருச் சொல்லை எடுத்துக்கொள்ளவும்; இதற்கு முன்னர் கூறியபடி '(புத்தக) அட்டை' மற்றும் '(மர) அட்டை' என்ற இரு பொருண்மைகள் உண்டு. இந்த இரண்டு பொருண்மைகளையும் பின்வரும் வாக்கியங்களில் உள்ள பொருத்தமான சேர்ந்துவருகைக் கட்டுப்பாட்டுப் பண்புக்கூறுகளை (selection restriction features) விளக்கி வேறுபடுத்தலாம்.

(19) புத்தகத்தின் அட்டை கிழிந்து விட்டது.

(20) அந்த அட்டை தரையில் ஊர்ந்து கொண்டிருந்தது.

கிழி என்ற வினை *அட்டை* போன்ற கிழியக்கூடிய பொருளை எழுவாயாக ஏற்கும். *ஊர்* என்ற வினை ஊர்ந்து செல்லக்கூடிய அட்டைப் பூச்சியை எழுவாயாக ஏற்கும்.

ஒப்புருச்சொன்மை மயக்கம் இலக்கண ஆய்வு அடிப்படையிலான வேறுபாடு அடிப்பையிலும் வெளிப்படும் (இலக்கணப்போலி). பின்வரும் எடுத்துக்காட்டுகள் இதை உணர்த்தும்.

(21) அவன் கடலை தின்று மகிழ்ந்தான்

(22) அவன் கடலைக் கண்டு மகிழ்ந்தான்.

21-ஆவது வாக்கியத்தில் *கடலை* என்பது *கடலை* என்ற தின்பண்டத்தைக் குறிப்பிடும் தனிச்சொல்லாகும். 9-ஆவது வாக்கியத்தில் *கடலை* என்பது *கடல்+ஐ* எனப் பகுத்தாய இயலும் சமுத்திரத்தைக் குறிப்பிடும் திரிபுற்ற சொல்லாகும்.

பின்வரும் வாக்கியங்களில் இரு சொற்களின் வேறுபட்ட இலக்கண வடிவுகள் ஒன்றுப்போல் அமைந்து ஒப்புருச்சொன்மை காட்டுவதன் காரணமாகப் பொருண்மை மயக்கம் காட்டுகின்றன.

(23) அவன் (துணி) நெய்தான்.

(24) அவன் நெய்தான் விருப்புகிறான்

23-ஆவது வாக்கியத்திலுள்ள *நெய்தான்* என்பதும் 24-ஆம் வாக்கியத்திலுள்ள *நெய்தான்* என்பதும் இருவேறு சொற்களின் இருவேறு இலக்கண வடிவங்கள். முதல் வாக்கியம் 'he weaved (a cloth)' என்றும் இரண்டாவது வாக்கியம் 'he likes only ghee' என்றும் பொருள்பட்டு வேறுபடும்.

வரலாற்று மொழியியல் அடிப்படையில் சொற்களின் திரிபு வடிவங்கள் வேற்பட்ட இலக்கணச் செயல்பாடுகளைப் பெற்று வேறுபட்ட சொல்வகைப்பாடுகளைக் குறித்துநிற்கும் அதாவது வேறுபட்ட பொருண்மைகளைக் குறித்துநிற்கும். இருப்பினும் அவை ஒரே வடிவில் இருப்பதால் அவற்றை ஒப்புருசொற்களின் வகையாகக் கருத இயலும். குறிப்பாக வினைகளின் திரிபுற்ற வடிவங்களிலிருந்து உருவாக்கப்பட்ட அல்லது மாற்றுச் செயல்பாடு பெற்ற முன்னுருபுகள் (post postions) இத்தகைய பொருண்மை மயக்கம் காட்டும். எடுத்துக்காட்டாக *இருந்து, பற்றி, குறித்து, ஒட்டி, கொண்டு, வைத்து, சுற்றி, நோக்கி, முந்தி, விட, கூட* என்ற முன்னுருபுகள் அவற்றின் மூல வினைகளின் திரிபுற்ற ஆனால் வகைப்பாடு (செயல்பாடு) மாறாத வடிவங்களுடன் பொருண்மை மயக்கம் காட்டும்.

25அ. அவன் வீட்டில் இருந்து வெளியேறினான். (முன்னுருபு)

25ஆ. அவன் வீட்டில் இருந்து வந்தான் (வினைத்திரிபு வடிவம்)

26அ. அவன் அவளைப் பற்றி பேசினான். (முன்னுருபு)

26ஆ. அவன் அவள் கையைப் பற்றி முத்தமிட்டான். (வினைத்திரிபு வடிவம்)

27அ. அவன் அவளைக் குறித்து பேசினான். (முன்னுருபு)

27ஆ. அவன் அவள் சொல்வதைக் குறித்து வந்தான். (வினைத்திரிபு வடிவம்)

28அ. அவன் அந்த தலைப்பை ஒட்டி பேசினான். (முன்னுருபு)

28ஆ. அவன் போஸ்டர் ஒட்டி பிழைக்கின்றான். (வினைத்திரிபு வடிவம்)

29அ. அவன் கத்தி கொண்டு அதை வெட்டினான். (முன்னுருபு)

29ஆ. அவன் கத்தியால் வெட்டிக் கொண்டு பேசினான். (வினைத்திரிபு வடிவம்)

30அ. அவன் கத்தி வைத்துப் பழம் வெட்டினான். (முன்னுருபு)

30ஆ. அவன் பணம் வைத்துக் கொண்டு சூதாடினான். (வினைத்திரிபு வடிவம்)

31.அ அவன் வீட்டைச் சுற்றி மரங்கள் நிற்கின்றன. (முன்னுருபு)

31ஆ. அவன் அவளையே சுற்றி வருகின்றான். (வினைத்திரிபு வடிவம்)

32அ. அவன் அவளை நோக்கி நடந்தான். (முன்னுருபு)

32ஆ. அவள் முகத்தை நோக்கிச் சிரித்தான். (வினைத்திரிபு வடிவம்)

33அ. அவன் அவளுக்கு முந்தி அங்கு வந்தான். (முன்னுருபு)

33ஆ. அவன் அவளை முந்தி நடந்துகொண்டிருந்தான். (வினைத்திரிபு வடிவம்)

34அ. அவன் அவளை விட நல்லவன். (முன்னுருபு)

34ஆ. அவன் அவளை விட விருப்பவில்லை. (வினைத்திரிபு வடிவம்)

35அ. அவன் அவள் கூட வந்தான். (முன்னுருபு)

35ஆ. அவனது மகிழ்ச்சி கூட ஆர்ப்பரித்தான்.

திரிபுற்ற *என்று* என்ற வடிவம் மூன்று வேற்பட்ட இலக்கணச் செயல்பாடுகளில் வந்து (இலக்கணப்போலி) பொருண்மை மயக்கம் காட்டும்.

36அ. அவள் நல்லவள் என்று நினைத்தேன். (என்று நிரப்பானாக வருகின்றது)

36ஆ. அவள் திடீர் என்று வந்தாள் (என்று வினையடையாக்கியாக வருகின்றது)

36இ. அவள் என்று வருகிறாள். (என்று காலத்தைக் குறிந்து வருகின்றது)

-*உம்* என்ற எதிர்கால வடிவம் வினை முற்றாகவும் பெயரெச்சாமாகவும் வரும்.

37அ. அது நாளை வரும்.

37ஆ. அது வரும் நாள் எனக்குத் தெரியாது.

-*அது* ஒட்டப்பட்ட வினைவடிவம் மூன்று வேறுபட்ட இலக்கணச் செயல்பாடுகளில் வந்து (இலக்கணப் போலி) பொருண்மை மயக்கம் காட்டும்.

38அ. அது நேற்று வந்தது (*வந்தது* வினைமுற்று வடிவம்)

38ஆ. அவன் வந்தது எனக்குத் தெரியாது (*வந்தது* வினையாலணையும் பெயர்)

38இ. அந்த செய்தித்தாள் நேற்று வந்தது (வந்தது அது மாற்றுப் பெயராக வினை)

தேர்வுக் கட்டுப்பாடு, சூழல் கட்டுப்பாடு, சூழல் அமைப்பு, சேர்ந்துவருகை என்பதன் அடிப்படையில் இத்தகைய சொற்களின் பொருண்மை மயக்கத்தை நீக்கலாம்.

2.3 மாற்றப் பொருண்மை மயக்கம்

வகைப்பாட்டு பொருண்மை மயக்கம், ஒப்புருச்சொன்மை, பல்பொருள் ஒருமொழியம் என்பன சொற்பொருண்மை மயக்கங்களுக்கு எடுத்துக்காட்டுகளாகும். அவை முதன்மையாக மூல மொழிப் பனுவல்களின் பகுப்பாய்வில் சிக்கலை விளைவிக்கின்றன. இயந்திர மொழிபெயர்ப்பில் மாற்றப் பொருண்மை மயக்கங்கள் (மொழிபெயர்ப்பு பொருண்மை மயக்கங்கள்) உள்ளன. ஒரு மூல மொழிச் சொல்லை இலக்கு மொழியில் பல சொற்களாலோ வெளிப்பாடுகளாலோ மொழி பெயர்க்க இயலும். மூல மொழிச் சொல் பொருண்மை மயக்கம் காட்டாது அல்லது பொருண்மை மயக்கம் இயல்மொழி பேசுபவர்களால் உணர்ந்து கொள்ளப்படாது இருக்கலாம். இச்சொல் பிற மொழி பேசுபவரின் புரிந்துகொள்ளலின் அடிப்படையில் தான் பொருண்மைமயக்கம் காட்டுகின்றது. எனவே இது மொழிபெயர்ப்பின் சிக்கலாகும்.

3. சொற்பொருண்மை மயக்கநீக்கம்: ஒரு கணினிசார் அணுகுமுறை

சொற்பொருண்மை மயக்கநீக்கம் என்பது ஒரு கணினிசார் வழி சூழலில் சொற்களின் பொருளைக் கண்டுகொள்ளும் ஒரு திறனாகும். சொற்பொருண்மை மயக்கநீக்கம் ஒரு செயற்கை அறிவு-முழுமைச் சிக்கலாகக் (AI-complete problem) கருதப்படுகின்றது; அதாவது, செயற்கை அறிவில் மிகக் கடினமான சிக்கல்களுக்குச் சமமான செயல்பாடாகும். இங்கே சொற்களின் மயக்கத்தை நீக்குவதற்கான செயலூக்கங்களும் இச்செயலின் விளக்கமும் தரப்படும். இங்கு கண்காணிக்கப்பட்ட, கண்காணிக்கப்படாத மற்றும் அறிவு அடிப்படையிலான அணுகுமுறைகள் சுருக்கமாகக் கூறப்படும்.

3.1 புற அறிவு மூலங்கள்

சொற்பொருண்மை மயக்கநீக்கத்தின் அடிப்படைக் கூறு/பகுதி அறிவு ஆகும். அறிவு மூலங்கள் சொற்களுடன் பொருண்மைகளைத் தொடர்புபடுத்தத் தேவையான தரவுகளைத் தருகின்றன. அவை புலக்குறிப்பு செய்யப்படாத (unlabeled) அல்லது அடையாளப்படுத்தப்பட்ட (annotated) பனுவல்களின் தரவுத்தொகுதிகளிலிருந்து (corpora of texts) இயந்திரத்தால் படிக்கவியலும் அகராதிகள் (machine readable dictionaries (MRDs)), பொருட்புல அகராதிகள் (thesauri), பொருள்விளக்கச் சொற்கோவை (glossaries), மூலப்பொருண்மையியல் ஆய்வுகள்/மெய்ப்பொருள் மூலாய்வு (ontologies) மற்றும் பிற என வேறுபடலாம்.

அமைப்பாக்கம் செய்யப்பட்ட மூலவளங்கள் (structured resources):

- பொருட்புல அகராதிகள்

    - இயந்திரத்தால் படிக்கவியலும் அகராதிகள்
    - மூலப்பொருண்மையியல் ஆய்வுகள் (Ontologies):

அமைப்பாக்கம் செய்யப்படாத மூலவளங்கள் (unstructured resources):

- தரவுத்தொகுதிகள் (corpora):
- கச்சாத் தரவுத்தொகுதிகள் (raw corpora)
- பொருண்மை அடையாளப்படுத்தப்பட்ட தரவுத்தொகுதிகள் (sense-annotated corpus)
- சேர்ந்துவருகை மூலவளம் (collocational resources)
- பிற மூலவளங்கள்
- சொல்வலை
    - பொருள்விளக்கம் (gloss)
    - சொல்லன்சார் மற்றும் பொருண்மைசார் உறவுகள் (Lexical and semantic relations)
        - எதிர்மொழியம் (antonymy):
        - சார்பு மொழியம் (Pertainymy):
        - பெயராக்கம் (Nominalization):
        - உள்ளடக்கு மொழியம் (hypernymy)
        - உள்ளடங்கு மொழியம் மற்றும் உட்படு மொழியம் (Hyponymy and troponymy):

3.2 கணினிசார் பொருண்மை மயக்க நீக்கத்தின் வகைகள்

கணினிசார் பொருண்மை மயக்கநீக்கதை மூன்றாகப் பிரிக்கலாம்.

1.கண்காணிக்கப்பட்ட பொருண்மை மயக்கநீக்கம்

2. கண்காணிக்கப்படாத பொருண்மை மயக்கநீக்கம்

3. அறிவு அடிப்படையிலான பொருண்மை மயக்கநீக்கம்

3.2.1. கண்காணிக்கப்பட்ட பொருண்மை மயக்கநீக்கம்

கண்காணிக்கப்பட்ட பொருண்மை மயக்கநீக்கம் பின்வருவனவற்றை உள்ளடக்கும்.

1. தீர்மான பட்டியல்கள்
2. தீர்மானக் கிளைகள்
3. நெய்வ் பெய்ஸ்
4. நரம்புசார் வலையமைப்பு
5. முன்மாதிரி அடிப்படையிலான அல்லது எடுத்துக்காட்டு அடிப்படையிலான கற்றல்

6. சப்போர்ட் வெக்டார் மெஷியன்கல்
7. ஒருங்கிணைந்த நெறிமுறைகள்
    - பெரும்பான்மை வாக்களிப்பு
    - நிகழ்தகைமைக் கலப்பு
    - தர அடிப்படையிலான ஒருங்கிணைப்பு
    - அடாபூஸ்ட்

8.மிகக்குறைவான மற்றும் பகுதி கண்காணிக்கப்பட்ட பொருண்மை மயக்கநீக்கம்

3.2.2 கண்காணிக்கப்படாத பொருண்மை மயக்கநீக்கம்

கண்காணிக்கப்படாத பொருண்மை மயக்கநீக்கம் பின்வருவனவற்றை உட்படுத்தும்.

1. சூழல் கொத்தாக்கம்
2. சொல் கொத்தாக்கம்
3. சேர்ந்துவருகை வரைபடங்கள்

3.2.3 அறிவு அடிப்படையிலான பொருன்மை மயக்கநீக்கம்

அறிவு அடிப்படையிலான பொருண்மை மயக்கநீக்கம் பின்வருவனவற்றை உட்படுத்தும்.

1. பொருண்மை மயக்கநீக்கங்களின் மேலுறல்
2. விருப்பத்தேர்வுகள்
3. அமைப்புமுறை அணுகுமுறைகள்
    - ஒற்றுமை அளவீடுகள்
    - வரைபட அடிப்படையிலான அணுகுமுறைகள்

3.2.4 பிற அணுகுமுறைகள்

4 தமிழில் சொற்பொருள் மயக்கநீக்கத்திற்கான நெறிமுறைகள்

ஒரு பனுவலில் உள்ள சொல் மயக்கங்களை நீக்க ஒரு பொருண்மை ஆய்வி தேவையாகும். இந்த ஒழுங்குமுறைத் தமிழை மூலமொழியாகக்கொண்ட இயந்திர மொழிபெயர்ப்புக்கும் தமிழ் தரவுகளின் அறிவாற்றல் உள்ள தகவல் மீட்பு ஒழுங்குமுறைக்கும் தேவையானதாகும். எடுத்துக்காட்டாகத் தமிழ்-ஆங்கில இயந்திர மொழிபெயர்ப்பு ஒழுங்குமுறை பின்வரும் வாக்கியத்தை உள்ளீடாகக் கொள்கின்றது என்று கொள்வோம்.

39. நீலகிரி காடுகளில் அரிய வகை விலங்குகள் காணப்படுகின்றன.

மேற்கண்ட வாக்கியத்தில் *விலங்குகள்* என்பது *விலங்கு* என்பதன் பன்மை வடிவமாகும். *விலங்கு* என்பதற்கு இரண்டு வேறுபட்ட அர்த்தங்கள் உள்ளன:

1. மிருகம்

2. கையில் போடும் விலங்கு

'விலங்கு' அர்த்தம் '(கை)விலங்கு' அர்த்தத்தைவிட அதிக நிகழ்வெண்ணில் வருகின்றது. ஆங்கிலத்தில் இந்த இரு அர்த்தங்களுக்கும் தனித்தனியான வடிவச் சொற்கள் உள்ளன: விலங்கு1 'animal', விலங்கு2 'hand-cufs'. எனவே பொருண்மையியல் ஆய்வி இப்பொருண்மை மயக்கத்தை நீக்க வேண்டும்; இதன் படி மேற்சொன்ன தமிழ் வாக்கியம் பின்வருமாறு மொழிபெயர்க்கப்பட வேண்டும்.

40. Rare animal species are found in Nilgiris forests.

கீழே எடுத்துக்காட்டாக இரு அணுகுமுறைகள் விவரிக்கப்பட்டுள்ளன.

4.1 கொத்தாக்கச் சேர்ந்துவருகை அணுகுமுறை

இவ்வாய்வில் பொருண்மை மயக்க நீக்கம் கொத்தாக்க உபாயம் (clustering technique) மற்றும் அறிவு மூலவளங்களாகச் சேர்ந்துவருகைகளையும் (collocations) வேற்றுமை குறியீடுகளையும் பயன்படுத்துவது அடிப்படையில் நடைமுறை படுத்தப்பட்டுள்ளது. ஒழுங்குமுறையின் செயல்பாடு படத்தில் தரப்பட்டுள்ள ஒழுங்குமுறை ஒழுக்குப் படத்தில் காட்டப்பட்டுள்ளது.

இவ்வணுகுமுறை இரண்டு கட்டங்களில் செயல்படுகின்றது. பயிற்சி கட்டத்தின் தொடக்கத்தில் ஒழுங்குமுறை பொருண்மை மயக்கம் உள்ள சொற்களைக் கண்டுபிடிக்கின்றது. வேறுபட்ட ஆய்வாளர்கள் வேறுபட்ட அளவுகளில் சூழல்களைப் (அதாவது வேறுபட்ட சாளர அளவுகளில் (different window size)) பயன்படுத்தியுள்ளனர். சாளர அளவு ஒழுங்குமுறையின் செயல்பாட்டைக் கணிசமான அளவில் பாதிக்கின்றது. ஒரு பெரிய சாளரம் சூழலுக்கு அதிக அளவில் தொடர்பில்லாத சொற்களைக் கொண்டுவருகின்றது; ஆனால் சிறிய சாளரம் சில முக்கியமான சேர்ந்துவருபவைகளை இழக்கின்றது. யரொவ்ஸ்கி (Yarowsky 1995) உட்பட பல சொற்பொருண்மை மயக்கநீக்கிகளில் காணப்படும் பொதுவான நடமுறையைப் பின்பற்றி இவ்வாய்வில் இருபது-சொல் சாளரம் பயன்படுத்தப்பட்டுள்ளது.

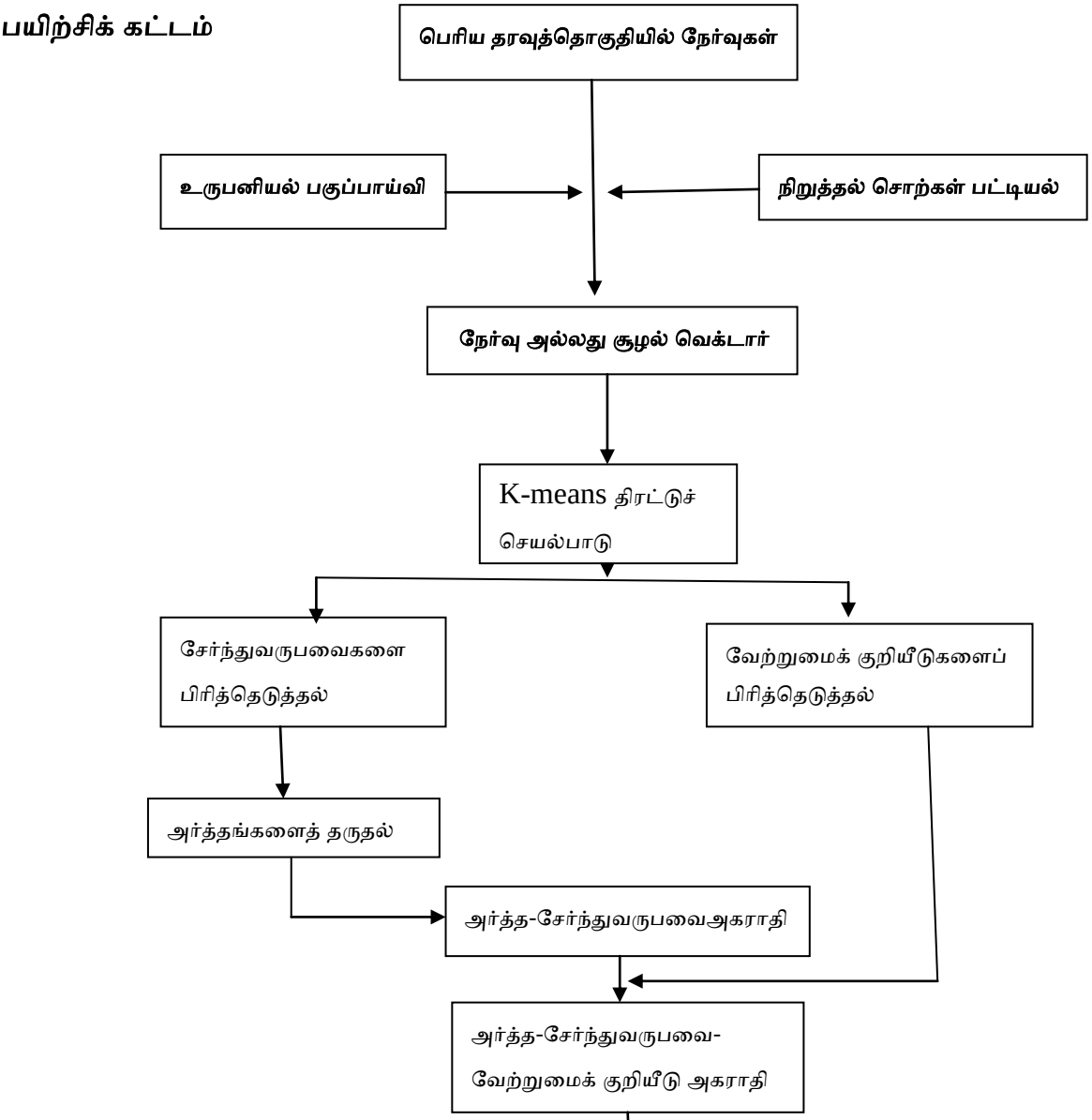
நிறுத்தல்-சொற்கள் (stop-words) என்று அழைக்கப்படுகின்ற அதிக நிகழ்வெண் உள்ள சொற்கள் சாளரத்திலிருந்து நீக்கப்படும். பின்னர் திரட்டுவதில் பயன்படுத்தப்படும் சூழல் சொற்களின் வேர்வடிவைப் பெற உருபனியல் அடிப்படையில் ஆயப்படும். ஒவ்வொரு சூழலும்

சூழல் அல்லது நிகழ்வு வெக்டராகப் உருப்படுத்தம் செய்யப்படும். சேர்ந்துவருபவைகள் தானியக்கமாக இந்தத் திரட்டுகளிலிருந்து சேகரிக்கப்படும் மற்றும் அர்த்தங்கள் மனித-அடையாளப்படுத்திகளால் (human-annotators) ஒதுக்கப்படும்/தரப்படும்; இதன்படி ஒரு அர்த்த-அகராதி உருவாக்கப்படும். வேற்றுமை-குறியீடுகள் ஒவ்வொரு திரட்டுக்கும் அடையாளம் காணப்படும் மற்றும் அர்த்த-சேர்ந்துவருபவை வேற்றுமை-குறியீடு அகராதியை (sense-collocation case-marker dictionary) உருவாக்க வேண்டி சேர்க்கப்படும்.

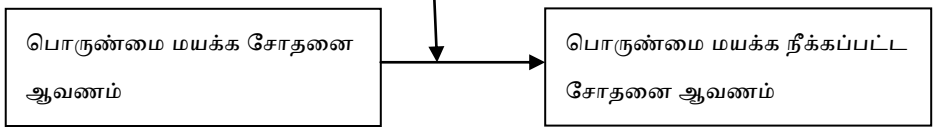
இரண்டாவது கட்டத்தில் ஒழுங்குமுறை பொருண்மை மயக்கநீக்கத்திற்குத் தயாராக இருக்கும். பொருண்மை மயக்கச் சொற்கள் கொண்ட எந்தப் பனுவலும் ஒழுங்குமுறைக்குத் தரப்படலாம். படத்தில் காட்டியுள்ளது போன்று ஒழுங்குமுறை பொருண்மை மயக்கச் சொற்களின் பார்க்கப்படாத நிகழ்வுகளை பொருண்மை மயக்கநீக்கம் செய்ய அர்த்த-சேர்ந்துவருகை வேற்றுமை-குறியீட்டு அகராதியைப் (Sense-Collocation-Case marker dictionary) பயன்படுத்துகின்றது. ஒவ்வொரு பொருண்மை மயக்கம் உள்ள சொற்களுக்கும் பயிற்சி ஒருநேரச் செயல்பாடாகும். ஒழுங்குமுறை ஒரு குறிப்பிட்ட சொல்லுக்காகப் பயிற்சி பெற்றுவிட்டால் அதை ஒரு தரப்பட்ட சூழலில் அச்சொல்லின் பொருண்மை மயக்கத்தை நீக்கப் பயன்படுத்த இயலும். இந்த நெறிமுறையில் சேர்துவருபவைகளும் வேற்றுமைக் குறியீடுகளும் ஒரு முக்கியமான பங்களிப்பு செய்கின்றது என்பதைக் காணலாம்.

ஒழுங்குமுறை ஒழுக்குப் படம்

பயிற்சிக் கட்டம்



சோதனைக் கட்டம்



4.1.1 சேர்ந்துவருபவைகளின் முக்கியத்துவம்

சேர்ந்துவருபவைகள் தந்துள்ள ஒரு வாக்கியத்தில் பொருண்மை மயக்கச் சொற்களின் அர்த்தத்தை உறுதியாக அனுமானிக்கும் அடுத்துவரும் சொற்கள் ஆகும். பொதுவாகச் சேர்ந்துவருபவை இரண்டு சொற்களுக்கு இடையில் அளக்க இயலும் இட-சிறப்பு உறவுகளைக் குறிப்பிடுகின்றது. சேர்ந்துவருபவைகள் பொருண்மை மயக்கமுள்ள சொற்களுக்குப் பொருண்மையியல் அடிப்படையில் மிக அண்மையில் வரும் சொற்களைப் பற்றிய குறியாக்கம் செய்யப்பட்ட தகவலாகும். எடுத்துக்காட்டாகப் பின்வருமாறு இரு அர்த்தங்கள் உள்ள plant என்ற சொல்லைக் கருத்தில் கொள்ளவும்.

plant1 – a living organism
plant2 – a manufacturing place

plant என்ற சொல்லின் சேர்ந்துவருபவைகள் பனுவலில் *plant* என்பதுடன் அடிக்கடிச் சேர்ந்துவரும் சொற்களை உட்படுத்தும். *plant* என்பதன் மேற்கண்ட அர்த்தங்களுக்கு எடுத்துக்காட்டான சேர்ந்துவருபவைகள் கீழே பட்டியலிடப்பட்டுள்ளன.

plant1 – growth, height, flower, fruit, species, leaves
plant2 – car, union, equipment, assembly, nuclear, job, worker

பெரும்பாலும் சேர்ந்துவருபவைகள் பொருண்மை மயக்கமுள்ள சொற்களிருந்து உள்ள தூரத்திற்கு உணர்வுள்ளவைகளாகும். பொருண்மை மயக்கச் சொற்களிலிருந்து குறைந்த தூரத்தில் உள்ள சொற்கள் வலுவாக அதன் அர்த்தத்தைச் சுட்டிக்காட்டும். தூரம் அதிகரிக்கும் போது பொருண்மைமயக்கச் சொற்களுடன் உள்ள அதன் சம்பந்தம் குறையும். தூரம் பொதுவாகச் சாளர அளவு என்று குறிப்பிடப்படும். இருப்பினும் சேர்ந்துவருவது என்று கருதுவதற்கு ஒரு சொல் பொருண்மைமயக்கம் உள்ள சொல்லிலிருந்து இருக்கவேண்டிய அறுதியான தூரம் இல்லை. சில வேளைகளில் சேந்துவருபவைகள் பொருண்மைமயக்கச் சொற்களிலிருந்து மிகுந்த தூரத்தில் வரலாம். தற்போதைய அணுகு முறையில் சேர்ந்துவரும் சொற்கள் யாவும் தானியக்கமாகத் திரட்டப்பட்ட நேர்வுகளிலிருந்து பிரித்தெடுக்கப்படும். இது விதை சேர்ந்துவருபவைகளைக் (seed collocations) கையால் தரும் தேவையை நீக்குகின்றது; ஒரு மொழியில் வரும் பொருண்மை மயக்கமுள்ள சொற்களைக் கருத்தில் கொள்ளும்போது இது சிரமமான வேலையாகும்.

4.1.2 வேற்றுமை குறியீடுகளின் முக்கியத்துவம்

ஆங்கிலத்தில் பெயருக்கும் வினைக்கும் உள்ள உறவுகள் *in, on. at, by, with* போன்ற முன்னுருபுகளால் வெளிப்படுத்தப்படும். தமிழில் இது பின்னுருபுகளால் வெளிப்படுத்தப்படும். *வர்ணம்* என்ற தமிழ்ச் சொல்லை எடுத்துக்கொள்வோம். இதற்கு குறைந்தது மூன்று வேறுபட்ட அர்த்தங்கள் உள்ளன.

1. நிறம்
2. இந்திய மரபு இசையில் ஒரு குறிப்பிட்ட பாடல் வகை
3. பண்டை இந்தியாவில் பிறப்பு அடிப்படையிலான பகுப்பு

இவ்வர்த்தங்கள் வாக்கியங்களில் வருகையில் வேற்றுமை உருபுகளின் தேர்வு சில வேளைகளில் இப்பொருண்மை வேறுபாடுகளை வெளிப்படுத்தலாம். *மாலை* என்ற சொல்லை எடுத்துக்கொள்வோம். இது மாலைக் காலத்தையும் கழுத்தில் அணியும் மாலையையும் குறிப்பிட்டுப் பொருண்மை மயக்கம் காட்டும். இவை பின்வருமாறு வாக்கியங்களில் பயன்படுத்தப்படலாம்.

மாலையில் பள்ளி ஆண்டுவிழா கொண்டாடப்பட்டது.

அவன் கழுத்தில் மாலை அணிவித்தார்கள்.

மேற்கண்ட வாக்கிய நேர்வுகளில் வேற்றுமை உருபுகள் அர்த்தங்களை வேறுபடுத்த உதவக்கூடும்.

4.1.3 கற்றல் கட்டம்

கொத்தாக்கல்/திரட்டல் செயற்பாங்கு (clustering process) இங்கு விளக்கப்பட்டுள்ளது.

சூழல்களும் சூழல் இடைவெளியும்

ஒரு ஆவணத்தில் சொற்பொருள் மயக்கமுள்ள சொல்லின் ஒவ்வொரு நேர்வும் அதன் சூழல் என்று அழைக்கப்படும். ஒரு பெரிய தரவுத்தொகுதியில் சொற்பொருண்மை மயக்கம் உள்ள சொல்லின் எல்லா சூழலகளும் S என்ற கருத்துச்சாயலால் குறிப்பிடப்பட்ட சூழல் இடைவெளியில் (context space) உருப்படுத்தம் செய்யப்படும். எடுத்துக்காட்டாக *மாலை* என்ற சொல்லைக் கருத்தில் கொள்வோம். 'மாலை (காலம்)' என்ற அர்த்தம் S1 என்றும் '(கழுத்திலிடும்) மாலை' என்ற அர்த்தம் S2 என்றும் கொள்வோம். மைசூரிலுள்ள இந்திய மொழிகளின் மைய நிறுவனத்தால் உருவாக்கப்பட்ட மூன்று மில்லியன் தமிழ்த் தரவுத்தொகுதியிலிருந்து பிரித்து எடுக்கப்பட்ட வாக்கியங்களில் வரும் *மாலை* என்பதை இரு அர்த்தங்களும் அடையாளப்படுத்தப்பட்டு கீழே தரப்பட்டுள்ளன.

தெரசாவின் அறப்பணிகளுக்கு நிகரில்லாத புகழ் மாலைகளாக அமைந்து மகிழ்வித்தன. S2

...தினந்தோறும் காலை மாலை ஆகிய இரு வேளைகளிலும் S1

காலையில் சூரியம் தோன்றிய நேரத்தில் இருந்து மாலையில் மறையும் நேரம் வரை S1

சூரியகாந்திப்பூ காலை, மாலை வேளைகளில் திசை S1

சுற்றுலாப் பேருந்து அடுத்த வெள்ளிக்கிழமை மாலை சென்னைக்குத் திரும்புகிறது S1

காலையில் 5 மணிக்கும், பிறகு 9:30 மணிக்கும், மலையில் 6:30 மணிக்கும் S1

அரசு தூக்கிலிட முடிவு செய்த நாளின் முதல் நாள் மலையே, தூக்கிலிட்டுக் S1

தத்தும் வீசிய துண்டறிக்கைகள் அன்று மாலையே "இந்துஸ்தான் டைம்ஸ்" S1

நடவடிக்கைகளின் காரணமாக நேற்று மாலையிலேயே வந்த செய்தியின்படி S1

"குரங்கு கையிலே பூமாலை" என்று அதைப்போல ஊராட்சி S2

சிலர், 'நான் மலையையே மலர் ஆக்குவேன் S2

அழகு, இளமை, கருவிழி, செல்லம், வாகனம், மாலை, கொடி முதலியவை S2

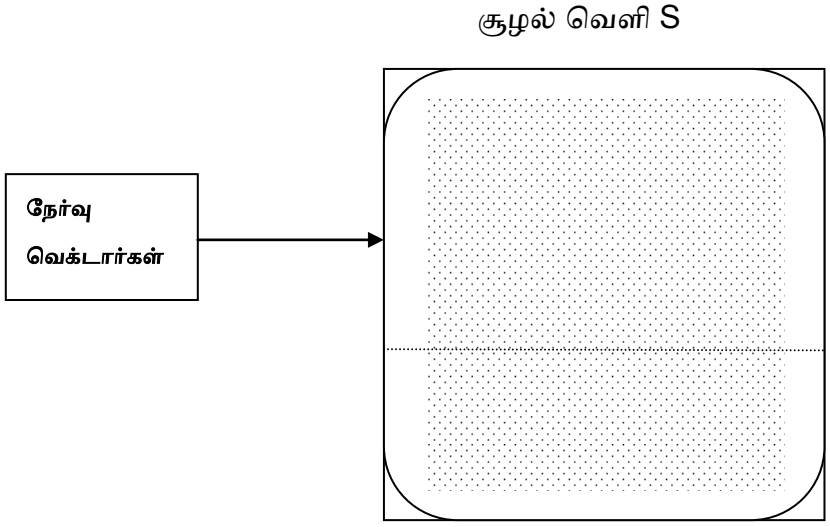
அல்லது திருநீறு காணப்படும் சபையோர் போடும் மாலைகளைப் கழற்றாமலே S2

நானும் அந்தச் செய்தியை நேற்று மாலை அமைச்சரவைக் கூட்டம் S1

உணவுப் பொருள்களை உட்கொள்ளும் பொழுதும், மாலையில் மேய்ச்சலுக்குப் பிறகும் S1

'மாலை (காலம்)' S1என்ற அர்த்தத்தில் வந்த எல்லா சூழல்களும் படத்தின் மேல்பாகத்திலும் '(கழுத்தில் அணியும்) மாலை' S2 என்ற அர்த்தத்தில் வந்த எல்லா சூழல்களும் படத்தின் கீழ்ப் பகுதியிலும் காட்டப்பட்டுள்ளன. பிரிக்கக்கூடிய புள்ளிக் கோடு இவ்விரு அர்த்தங்களின் சூழல் வெளியைக் காட்டுவதற்காக இடப்பட்டுள்ளது. கொத்தாக்கல்/திரட்டுச் செயற்பாங்குகள் வேறுபட்ட சுழல்களின் உண்மையான பிரிவை எய்துகின்றது.

சொற்பொருண்மை மயக்கம் உள்ள சொல்லின் ஒவ்வொரு சூழலும் சூழல் வெக்டார்களாகக் கையாள இயலும்; அவை சூழல் வெளியில் உருப்படுத்தம் செய்யப்படும். சூழல் வெக்டாரின் தனிமங்கள் பொருள்மயக்கம் உள்ள சொல்லின் எல்லா சூழல்களிலும் தோன்றும் எல்லா சொற்களையும் உட்படுத்தும். சூழல்களை சூழல் வெக்டார்களாக உருப்படுத்தம் செய்யும் நெறிமுறை ஒரு எடுத்துக்காட்டால் கீழே காட்டப்பட்டுள்ளது.



இப்போது பொருள்டக்கச் சொற்களை அடையாளம் காண்பதன் செயற்பாங்கை விளக்க மூன்று சூழல்களைக் கருத்தில் கொள்ளவும்.

C1 – மாலையில் நடைபெற்ற விழாவில் மாணவர்களுக்குப் பரிசு வழங்கப்பட்டது

C2 – வரவேற்பு விழா மாலை சரியாக ஆறு மணிக்குத் துவங்கியது.

C3 – ரோஜாப்பூ மாலை இருபது ரூபாயில் கிடைக்கிறது

சூழல் சாளரத்தில் வரும் எல்லாச் சொற்களும் உருபனியல் அடிப்படையில் ஆயப்பட்டது மற்றும் விளையும் வேர்வடிவுகள் திறனுள்ள பொருளட்டக்கச் சொற்களாகக் கருதப்படும். இவை நிறுத்தல் சொற்களை நீக்கி மேலும் குறைக்கப்படும். C1-க்கு ஆறு பொருளடக்கச் சொற்களைக் கண்டுபிடிக்க இயலும்.

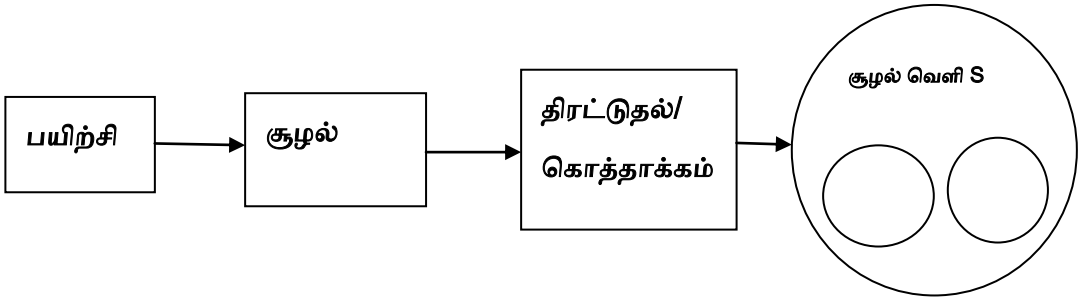
மாலை, நடைபெற்ற, விழா, மாணவர், பரிசு, வழங்கு

இதே வழியில் தரவுத்தொகுதியில் வரும் எல்லா நேர்வுகளுக்கும் பொருளடக்கச் சொற்களைக் கண்டுபிடிக்க இயலும். இவ்வாறு நமது குறைந்த மூன்று நேர்வுகளுக்கு 5 பொருளடக்கச் சொற்களைக் கண்டுபிடிக்க இயலும்.

மலை, நடைபெற்ற, விழா, மாணவர், பரிசு, வழங்கு, வரவேற்பு, ஆறு, மணி, துவங்கு, ரோஜாப்பூ, இருபது, ருபாய், கிடைக்கிறது

ஒவ்வொரும் நேர்வும் ஒரு சூழல் வெக்ட்டராக இப்போது உருப்படுத்தம் செய்யப்படும்; இதைச் சூழல் வெளியில் ஒரு தனிப்பட்ட இடத்தில் காட்ட இயலும். சூழல் வெக்டாரின் நீளம் தரவுத்தொகுதியில் எல்லா நேர்வுகளுக்கும் அடையாளம் கண்டுபிடிக்கக்கூடிய பொருளடக்கச் சொற்களின் எண்ணிக்கை ஆகும். இவ்வாறு எடுத்துக்காட்டாக *C1*, *C2* மற்றும் *C3* முன்று சூழல்களுக்கு ஒவ்வொரு சூழல் வெக்டாரின் நீளம் 14 ஆகும். சூழல் வெக்டாரில் உள்ள ஒவ்வொரு தனிமமும் அந்த குறிப்பிட்ட சொல் அச்சூழலில் இருக்கிறதா அல்லது இல்லையா என்பதைப் பொறுத்து பூஜியத்தையோஅல்லது ஒன்றையோ எடுக்கும். திரட்டும் வழிமுறை வரைவு பொருண்மை மயக்கம் உள்ள சொல்லின் நேர்வுகளை வேறுபட்ட திரட்டுகளாகக் குழுமும்; ஒவ்வொரு கொத்தும்/திரட்டும் சூழல் வெளியில் ஒரு குறிப்பிட்ட இடத்தில் செறிவாக இருக்கும். கொத்தாக்கத்தின்/திரட்டுவதன் அடிப்படைக் கருத்து ஒரு கொத்தில்/திரட்டில் சூழல் அடிப்படையில் ஒற்றுமையுள்ள நேர்வுகளைக் குழுமுவதாகும். பின்வரும் படம் இவ்வணுகுமுறையில் கொத்தாக்கத்தின்/திரட்டும் செயற்பாங்கின் செயல்பாட்டைத் தெளிவாகக் காட்டுகின்றது. இது பொருண்மை மயக்கநீக்கதிற்கு வேண்டி இங்குப் பயன்படுத்தப்பட்டுள்ள திரட்டும் உபாயத்தை எடுத்துக்காட்டுகின்றது.

தரவுத்தொகுதியிலிருந்து பொருண்மை மயக்கச்சொற்களின் எல்லாச் சூழல்களும் முதலில் சூழல் வெக்டார்களாக உருப்படுத்தம் செய்யப்படும். பின்னர் சூழல் வெக்டார்கள் எல்லாம் K-மீன்ஸ் வழிமுறைவரைவைப் பயன்படுத்தித் திரட்டப்படும்.

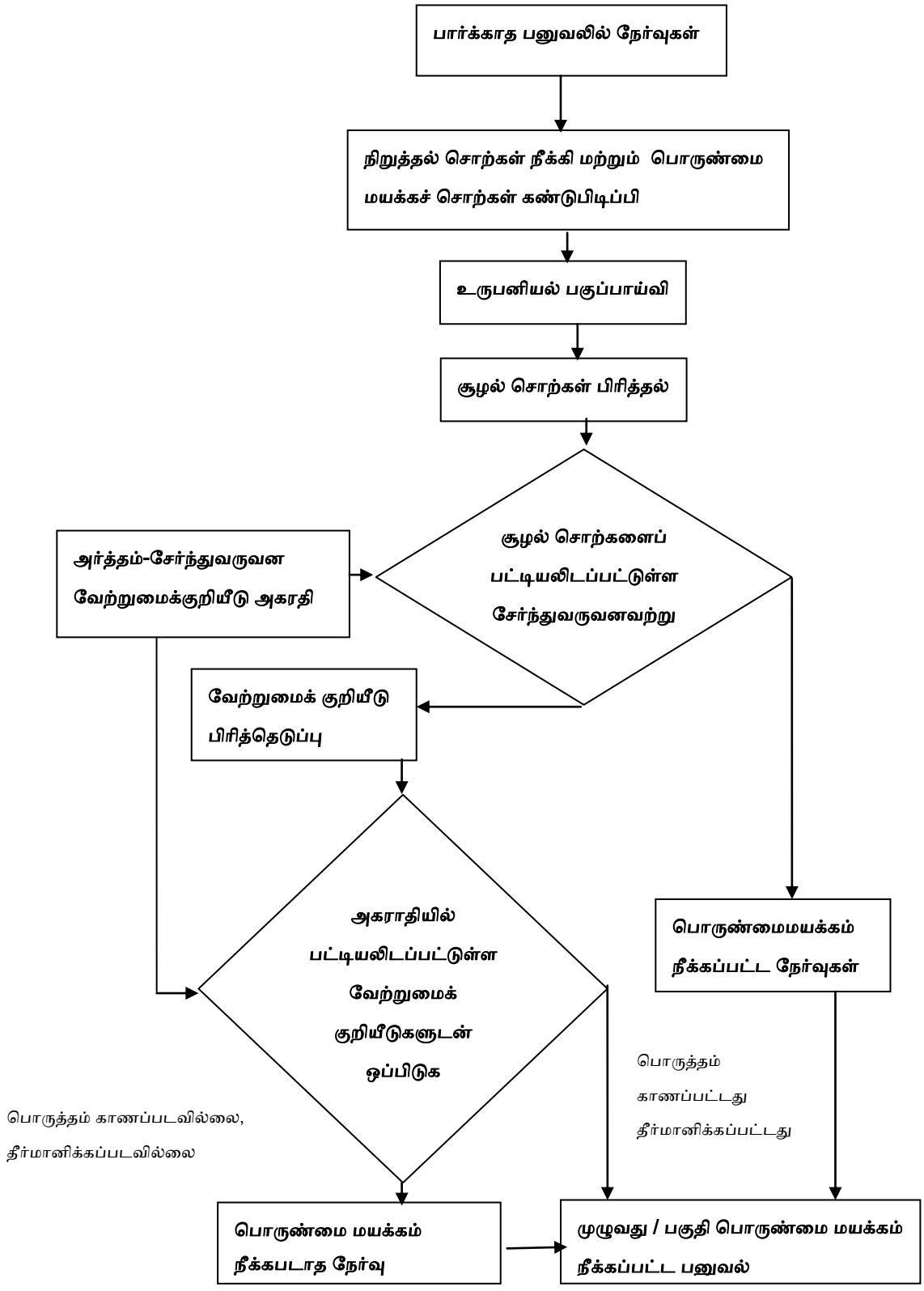


கொத்தின்/திரட்டின் இறுதியில் சூழல் அடிப்படையில் ஒத்திருக்கின்ற கொத்துக்/திரட்டுக் குழுக்கள் பெறப்படும். கொத்துக்களிலிருந்து/திரட்டுகளிலிருந்து சேர்ந்துவருபவைகளைப் பிரித்தெடுத்தபின் அடுத்த நடவடிக்கை அர்த்தம் அடையாளப்படுத்துவது ஆகும். இங்கு மனித அடையாளப்படுத்துபவர்கள் உயர்ந்த பத்து சேர்ந்துவருகைகளும் பொருண்மை மயக்கச் சொற்களின் அர்த்தங்களும் தரப்படுவார்கள். சொற்கள் வேறுபட்ட குழுமங்களாகப் பிரிக்கப்படும் மற்றும் ஒவ்வொரு குழுமமும் வேறுபட்ட அடையாளப்படுத்துபவரிடம் கொடுக்கப்படும். ஒவ்வொரு சேர்ந்துவரும் குழுமத்திற்கும் கூடுதல் பொருத்தமான அர்த்தத்தைத் தொடர்புபடுத்தி ஒரு அர்த்தம்-சேர்ந்துவருகை அகராதி உருவாக்குவதாகுவது அடையாளப்படுத்துபவர்களின் வேலையாகும்.

சொல்	அர்த்தம்	சேர்த்துவரும் சொல்
மாலை	மாலைப் பொழுது	காலை, பகல், சூரியன், மறை, நேரம், நாள், மாதம், மதியம், இரவு
மாலை	(கழுத்திலிடும்) மாலை	பூ, மணம், ரோஜா, மல்லிகை, மலர், பூஜை, மகளிர், கொன்றை
நூல்	நூல் (இழை)	இழை, தறி, சாயம், நெய், கதர், நெசவு, ஊடு, பாவு, துணி, ஊசி
நூல்	புத்தகம்	ஆசிரியர், பதிப்பகம், அச்சு, பக்கம், தாள்
விலங்கு	மிருகம்	காடு, சிங்கம், புலி, தோல், யானை, தந்தம், நாய், பூனை
விலங்கு	(கை) விலங்கு	குற்றவாளி, கோர்ட், போலீஸ், கைதி, விசாரணை, தண்டனை

4.1.4 பரிசோதனைக் கட்டம்

இப்போது ஒழுங்குமுறை பயன்பத்துபவர் பயன்படுத்துவதற்குத் தயாராக உள்ளது. இதில் ஒருவர் பொருண்மை மயக்கச் சொற்கள் உள்ள தமிழ்ப் பனுவலை உள்ளீடாகக் கொடுக்க இயலும் மற்றும் முழுவதும் அல்லது பகுதி பொருண்மை மயக்கம் நீக்கப்பட்ட பனுவலை வெளியீடாகப் பெற இயலும். பரிசோதனைக் கட்டத்தில் அர்த்தம்-சேர்ந்துவருகை வேற்றுமை குறியீட்டு அகராதி பார்க்கப்படாத நேர்வுகளைப் பொருண்மை மயக்கம் நீக்குவதற்குப் பயன்படுத்தப்படுகின்றது. பரிசோதனைக் கட்டத்தில் கண்டுபிடிக்கப்படும் சேர்ந்துவருவனவும் வேற்றுமைக் குறியீடுகளும் தரப்பட்ட பனுவலில் உள்ள பொருண்மை மயக்கங்களை நீக்கப் பயன்படுத்தப்படுகின்றது. இருப்பினும் வழிமுறை வரைவு தரப்பட்ட பனுவலில் தோன்றும் பொருண்மை மயக்கமுள்ள சொற்களுக்கும் முன்னரே பயிற்சிசெய்யப்பட்டிருக்கவேண்டும்.


பார்க்காத பனுவலில் நேர்வுகள்
நிறுத்தல் சொற்கள் நீக்கி மற்றும் பொருண்மை மயக்கச் சொற்கள் கண்டுபிடிப்பி
உருபனியல் பகுப்பாய்வி
சூழல் சொற்கள் பிரித்தல்
சூழல் சொற்களைப் பட்டியலிடப்பட்டுள்ள சேர்ந்துவருவனவற்று
அர்த்தம்-சேர்ந்துவருவன வேற்றுமைக்குறியீடு அகரதி
வேற்றுமைக் குறியீடு பிரித்தெடுப்பு
அகராதியில் பட்டியலிடப்பட்டுள்ள வேற்றுமைக் குறியீடுகளுடன் ஒப்பிடுக
பொருண்மைமயக்கம் நீக்கப்பட்ட நேர்வுகள்
பொருத்தம் காணப்பட்டது தீர்மானிக்கப்பட்டது
பொருத்தம் காணப்படவில்லை, தீர்மானிக்கப்படவில்லை
பொருண்மை மயக்கம் நீக்கபடாத நேர்வு
முழுவது / பகுதி பொருண்மை மயக்கம் நீக்கப்பட்ட பனுவல்

4.2 கருத்துரு வரைபட அகராதி அடிப்படையில் சொற்பொருண்மை மயக்க நீக்கம்

இவ்வாய்வில் சொற்பொருண்மை மயக்கநீக்கத்திற்கு வேண்டி கருத்துரு வரைபட அகராதி/ கருத்துரு அகராதி உருவாக்கப்பட்டுப் பயன்படுத்தப்பட்டுகின்றது. கருத்துரு வரைபட அகராதியின் உருவக்கம் பற்றியும் அதைப் பயன்படுத்திச் சொற்பொருண்மை மயக்க நீக்கம் எவ்வாறு மேற்கொள்ளப்படுகின்றது என்றும் இங்கு விரிவாக விளக்கபட்டுள்ளது. இதற்கு அடிப்படையாக அமையும் சொற்பொருண்மை ஆய்வு (lexical semantics study) பற்றி முன்னர் விளக்கப்பட்டுள்ளது. இது அறிவு அடிப்படையிலான நெறிமுறையை/அணுகுமுறையைச் சாரும். இந்நெறிமுறையில் கருத்துரு வரைபட அகராதி சொற்களின் பொருண்மை மயக்கநீக்கத்திற்குப் பயன்படுத்தப்படுகின்றது. இது ஒரு மாதிரி (model) நிலையிலேயே அமைந்துள்ளது. பொருண்மை மயக்கம் தருகின்ற சில சொற்கள் தேர்ந்தெடுக்கப்பட்டுத் தரவுத்தொகுதியிலிருந்து அச்சொற்கள் வரும் வாக்கியங்கள் பிரித்தெடுக்கப்பட்டு பொருண்மை மயக்கநீக்கம் செய்ய முயற்சி எடுக்கப்பட்டுள்ளது. இதற்கு என்று அமிர்தாப் பல்கலைக்கழகத்தின் தரவுத்தொகுதி பயன்படுத்தப்பட்டுள்ளது முழுச் சொற்றொகைக்கும் தற்போது இதைப் பயன்படுத்த இயலாது. இருப்பினும் ஒரு முழுமையான கருத்துரு வரைபட அகராதி உருவாக்கப்பட்டால் இது ஓரளவுக்குச் சாத்தியமாகும்.

இது இருகட்டங்களைக் கொண்டது: (1) கற்றல் கட்டம் (2) பரிசோதனைக் கட்டம். முதல் கட்டத்தில் கருத்துரு வரைபட அகராதியைப் பயன்படுத்தி தரவுத்தொகுதியில் வரும் பொருண்மை மயக்கமுள்ள சொற்கள் வேறுபட்ட கருத்துருக்களுக்காக (அதாவது பொருண்மைகளுக்காக/அர்த்தங்களுக்காக) அடையாளப் படுத்தப்படும். இரண்டாம் கட்டத்தில் அடையாளப்படுத்தப்படாத தரவுத்தொகுதியில் உள்ள செற்களின் பொருண்மை மயக்கங்களை நீக்கப் பரிசோதனை செய்யபடும்.

4.2.1 கருத்துரு வரைபடம்

செயற்கை அறிவு வெளிப்படுத்தத்திற்குப் பயன்படுத்தப்படும் கருத்துரு சார்ந்த அமைப்பு அடிப்படையிலான கோட்பாடு சௌவா (Sowa, 1984) என்பவரால் விளக்கப்பட்டுள்ளது. இது மூளையிலும் இயந்திரத்திலும் (கணிப்பொறியிலும்) தகவலை ஆய்வதற்கான சாத்தியம்

அடிப்படையில் அமைந்தது. கருத்துரு வரைபடம் நபர்களுக்கும் (persons) பொருளுக்கும் (things) நிகழ்வுகளுக்கும் (events) இடையிலுள்ள உறவுகளைக் கூறும் தர்க்க வடிவங்களாகும். சாம்ஸ்கி மாற்றிலக்கணம் அறிதிறனின் ஒரு அருவக் கோட்பாடு என்றும் அது செயல் திறனின் கோட்பாடாகக் கருதப்படக்கூடாது என்பார். ஆனால் செயற்கை அறிதிறனுக்கு வழியமைப்பாளர்கள் மக்களுக்கும் கணிப்பொறிக்கும் இடையில் நடக்கும் கருத்துப்பரிமாற்றத்திற்கு உதவிபுரியும் செயல்திறன் கோட்பாட்டை வேண்டுவர். செயற்கை அறிதிறனில் கருத்துரு வரைபடம் பரவலாகப் பயன்படுத்தப்படுகின்றது. பின்வரும் கருத்துரு வரைபடம் எடுத்துக்காட்டாகத் தரப்பட்டுள்ளது.

பையன் நாயைக் கடித்தான்.

பின்வரும் கருத்துரு வரைபடம் இந்த வாக்கியத்தின் அடிப்படைப் பொருளை உருப்படுத்தம் செய்யும்.

[பையன்]← (செயலி) ←[கடி] →(செ.பொ) →[நாய்]

பெட்டிகள் கருத்துருக்கள் (செவ்வக அடைப்புக் குறிகளுக்குள் தரப்பட்டுள்ளது) என்றும் வட்டங்கள் (பிறை அடைப்புக்குறிக்குள் தரப்பட்டுள்ளது) கருத்துரு உறவுகள் என்றும் அழைக்கப்படும். கருத்துரு எந்த ஒரு மொழியில் வரும் எந்த இருப்புப் பொருளையும் (அதாவது செயலையோ நிலையையோ) உருப்படுத்தம் செய்யும். கருத்துரு உறவுகள் ஒவ்வொரு இருப்புப் பொருளும் பங்காற்றுகிற பாத்திரத்தைக் குறிக்கும். மேற்கண்ட எடுத்துக்காட்டில் [பையன்] என்பது [கடி] என்ற செயலின் (செயலி)-ஆகச் செயல்படும் கருத்துருவை உருப்படுத்தம் செய்யும். [பையன்] என்ற கருத்துரு [கடி] என்பதன் (செயப்படுபொருள்)-ஆகப் பங்கெடுக்கும் இருப்புப் பொருளை உருப்படுத்தம் செய்யும். கருத்துரு வரைபடத்தார் எல்லா பொருண்மை குறித்த செய்திகளையும் ஒரு அடி அமைப்பில் இருப்பதாகக் கொள்ளாமல் ஆறு வேறுபட்ட வகையிலான செய்திகளிலிருந்து வாக்கியங்கள் ஆக்கப்படுவதாகக் கொள்கின்றனர். அவையாவன: இறந்தகாலமும் உள்ளது கிளத்தல் வினை நோக்கும் (indicative mood) செயல் குறிப்பிட்ட நேரத்திற்கு முன்னர் நிகழ்ந்ததைக் குறிப்பிடும். *பையன்* ஒரு சூழலில் ஒரு குறிப்பிட்ட நபரின் முன்வரு கிளவியாக வருகிறது என்பது முற்கோளாகும். *அவன்* சிறிது முன்னர் குறிப்பிடப்பட்டிருக்கலாம் அல்லது அந்த காட்சியில் இருக்கலாம். *நாய்* கடிப்பதைப் பற்றிய புதிய செய்தி கவனக்குவிப்பாகும். பையன் முன்னர் கூறப்பட்ட ஒருவரைக் குறிப்பதன் காரணமாக ஒரு உடனிலைக் குறிப்பு முன்னர் வரும் பையனை *பையன்* என்ற கருத்துருவுடன் தொடர்புபடுத்துகின்றது. ஆறாவதான உணர்ச்சி உணர்பொருள்கள் வாக்கியத்தில் உள்ள

வார்த்தைகளாலோ வரைபடத்திலுள்ள கருத்துருக்களாலோ நேரடியாக வெளிப்படுத்தப்படவில்லை. ஆனால் அது மக்களுக்குள் வேறுபடும் முந்தைய அனுபவங்களின் சேர்க்கைகளால் நிறுவப்பெறும்.

4.2.2 கருத்துரு வரைபட அகராதியைப் பயன்படுத்தி பொருண்மை மயக்கநீக்கம் செய்தல்

இவ்வாய்வில் மொழியியல் தகவல்களைப் பயன்படுத்தி சொற்பொருண்மை மயக்கநீக்க செய்வதற்கான முயற்சி எடுக்கப்பட்டுள்ளது. இதற்கென்று கருத்துரு வரைபட அகராதி மாதிரியாக உருவாக்கப்பட்டுள்ளது. சொற்கள் அவை உணர்த்தும் பொருண்மை அடிப்படையில் கருத்துருக்களாகப் பிரிக்கப்படும். ஒரு கருத்துரு வரைபடத்தில் அக்கருத்துரு குறித்த எல்லா தகல்களும் வெளிப்படுத்தப்பட்டிருக்கும். இதற்கு புஸ்தெஜொவ்ஸ்கியின் ஆக்கமுறை அகராதி கோட்பாட்டிலிருந்து சில செய்திகள் இங்கு பயன்படுத்தப்பட்டுள்ளன. கருத்துருவரைபட அகராதி கருத்துருக்களுக்கான விளக்கம் மட்டுமன்றி அக்கருத்துகுறித்த எல்லாச் செய்திகளையும் தரும். ஒரு விதத்தில் கூறினால் இது ஒரு களஞ்சியச் செய்திகளை உள்ளடக்கும்.

புஸ்தொஸ்கி கோட்பாடு அடிபடையில் ஒரு சொலுக்குச் பின்வரும் சொற்பொருண்மை தகவல்கள் தரப்படவேண்டும்.

- உறுப்பு உறவு (constitutive relation): ஒரு பொருளுக்கும் (object) அதன் உறுப்புப் பாகங்களுக்கும் (constituent parts) இடையில் உள்ள உறவு.
- வடிவ/முறை உறவு (formal relation): பொருளை (object) பொரிய சொற்களத்திற்குள் வேறுபடுத்தம் செய்கிறது.
- நோக்க உறவு (telic relation): பொருளின் தேவையையும் செயல்பாட்டையும் தருகிறது.
- செயலி உறவு (agentive relation): பொருளின் மூலம் அல்லது அதைக் கொண்டு வருகையில் உட்படும் காரணிகளைத் தருகின்றது.

எடுத்துக்காட்டாகப் புதினம் என்ற கருத்துருவை எடுத்துக்கொள்வோம். அதன்
சொற்பொருண்மைத் தகவல்கள் பின்வருவனவற்றை உள்ளடக்கும்

புதினம்	உறுப்புப்பங்களிப்பு=கதை
....	வடிவப் பங்களிப்பு=புத்தகம்
குண அமைப்பு =	நோக்கப் பங்களிப்பு=படிப்பது
	செயலிப் பங்களிப்பு=எழுவது

புத்தகம் என்ற கருத்துருவை எடுத்துக்கொண்டால் அது குறைந்தது பின்வரும் செய்திகளால் விளக்கப்படவேண்டும்.

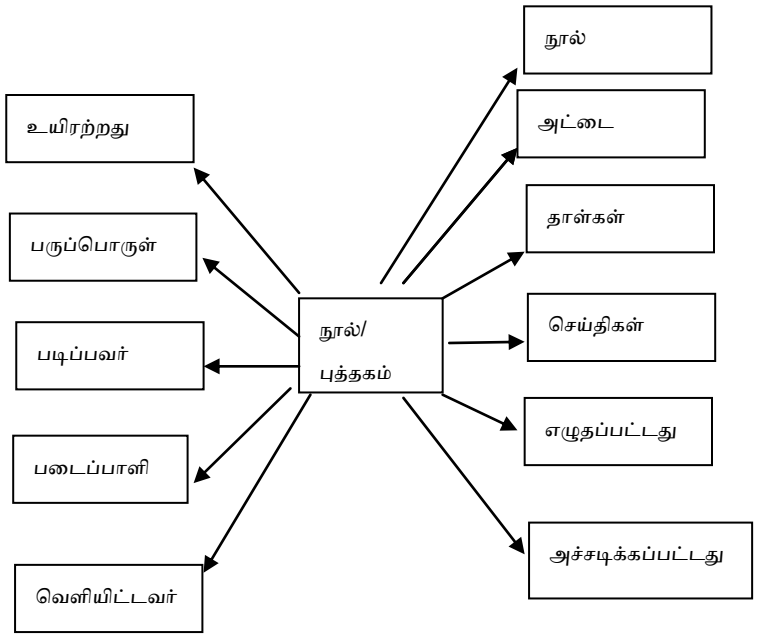
அதன் உறுப்புப் பங்களிப்பு = அது படிப்பதற்கான ஒரு பொருள்

அதன் வடிவப் பங்களிப்பு = அது தாள்களால் ஆனது, அட்டை போடப்பட்டது, படிப்பதற்கான செய்திகள் அடங்கியது

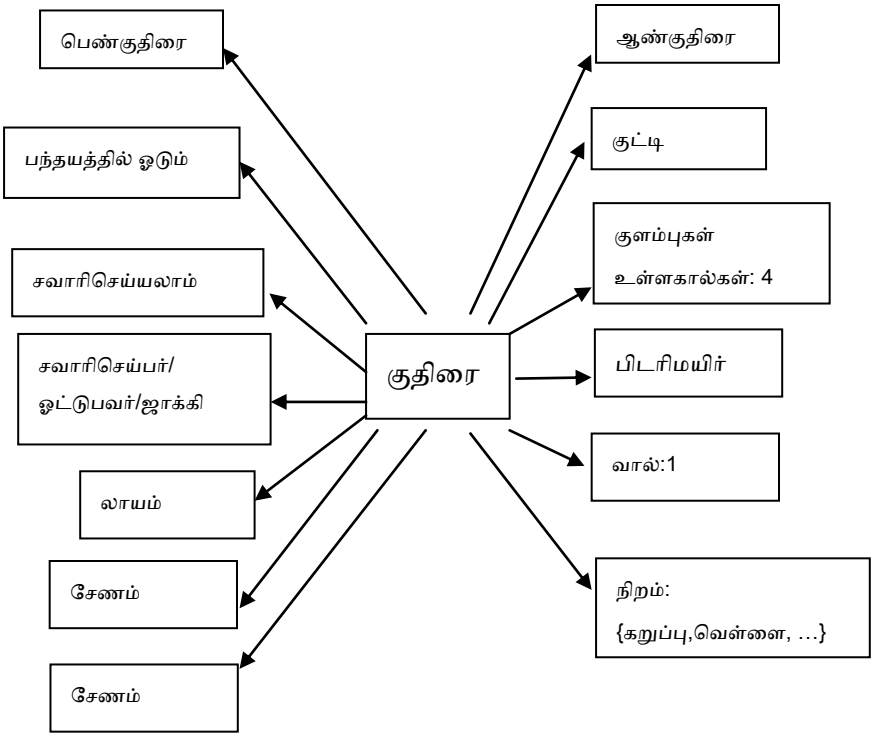
நோக்கப் பங்களிப்பு = படிப்பது

செயலி பங்களிப்பு = எழுதுவது

இதன் படி *புத்தகம்* என்பது பின்வரும் விளக்கத்தைப் பெறும்: படிப்பதற்குப் பயன்படும் எழுதப்பட்ட, வெளியிடப்பட்ட தாளகளால் ஆன, அட்டை போடப்பட்ட, படிப்பதற்கான செய்திகள் அடங்கிய ஒரு திடப்பொருள். இதைப் பின்வருமாறு கருத்துரு வரைபடமாகத் தரல்லாம். கருத்துருக்களுக்கு இடையே உள்ள உறவுகள்தரப்படவில்லை



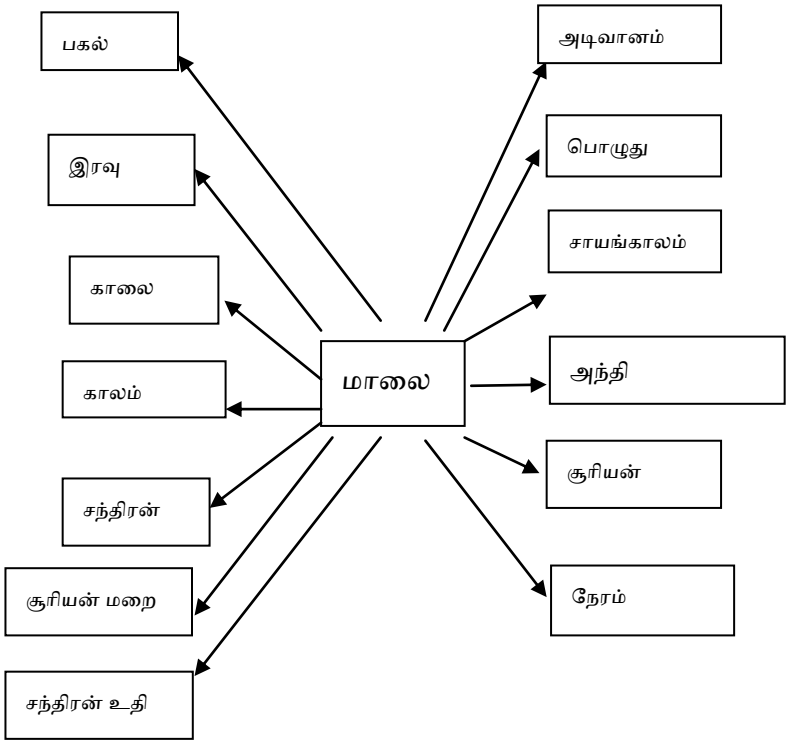
இதுபோல் *குதிரை* என்ற கருத்துருவை எடுத்துக்கொண்டால் அது பின்வருமாறு தகவல்களை வேண்டும்



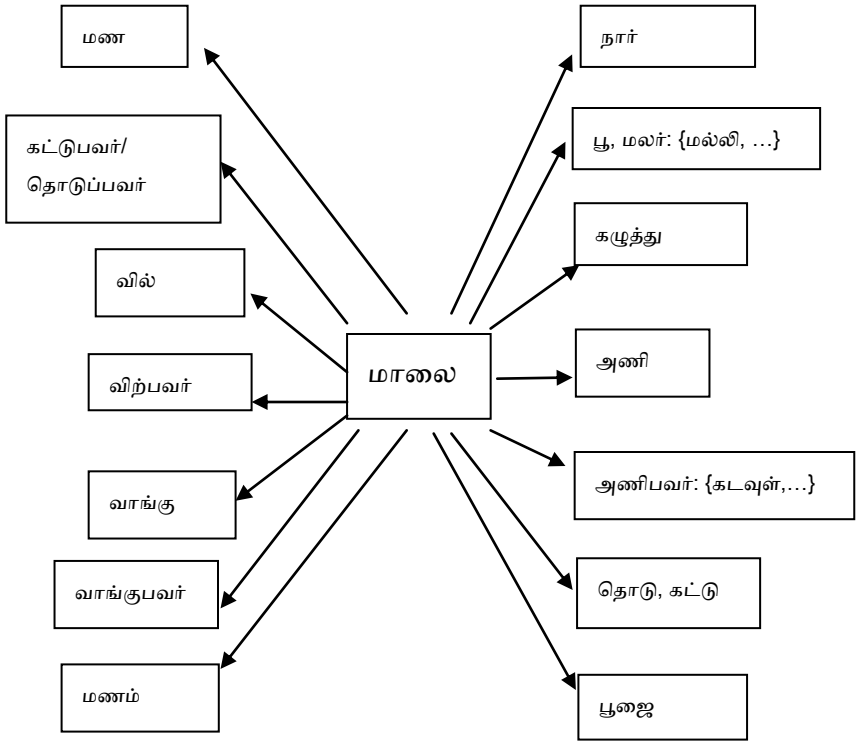
இங்கு முன்மொழிகின்ற கருத்துரு வரைபட அகராதி குறைந்தது மேற்கண்ட தகவல்களைக் கொண்டிருக்க வேண்டும்.

இக்கருத்துரு வரைபட அகராதியைக் கொண்டு முந்தைய அணுகுமுறையில் சுட்டிக்காட்டிய சொற்பொருள் மயக்கம் உள்ள சொற்களின் பொருண்மை மயக்கத்தையும் பிற சொற்களின் பொருள்மயக்கங்களையும் நீக்கலாம். எடுத்துக்காட்டாக *மாலை1* (பொழுது), (கழுத்தில் போடும்) *மாலை2* என்ற கருத்துருவின் கருத்துரு வரைபட அகராதி ஓரளவுகுப் பின்வருமாறு அமையும்.

மாலை1


மாலை
பகல்
இரவு
காலை
காலம்
சந்திரன்
சூரியன் மறை
சந்திரன் உதி
அடிவானம்
பொழுது
சாயங்காலம்
அந்தி
சூரியன்
நேரம்

மாலை2

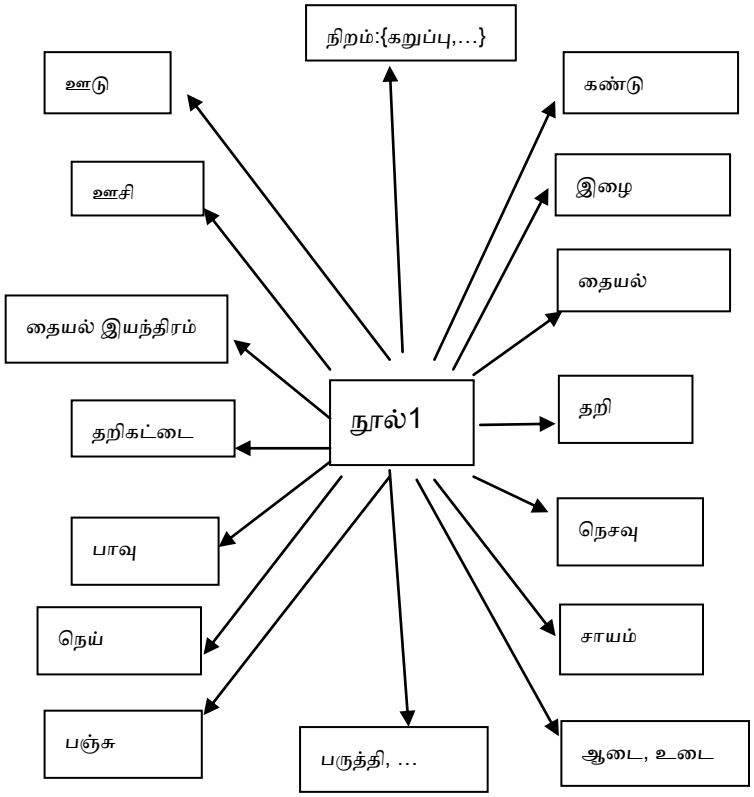

மாலை
மண
கட்டுபவர்/
தொடுப்பவர்
வில்
விற்பவர்
வாங்கு
வாங்குபவர்
மணம்
நார்
பூ, மலர்: {மல்லி, …}
கழுத்து
அணி
அணிபவர்: {கடவுள்,…}
தொடு, கட்டு
பூஜை

மேற்கண்ட *மாலை1*, *மாலை2* என்ற கருத்துரு வரைபட அகராதி விளக்கம் அடிப்படையில் தரவை பின்வருமாறு பொருண்மை மயக்க நீக்கம் செயலாம்.

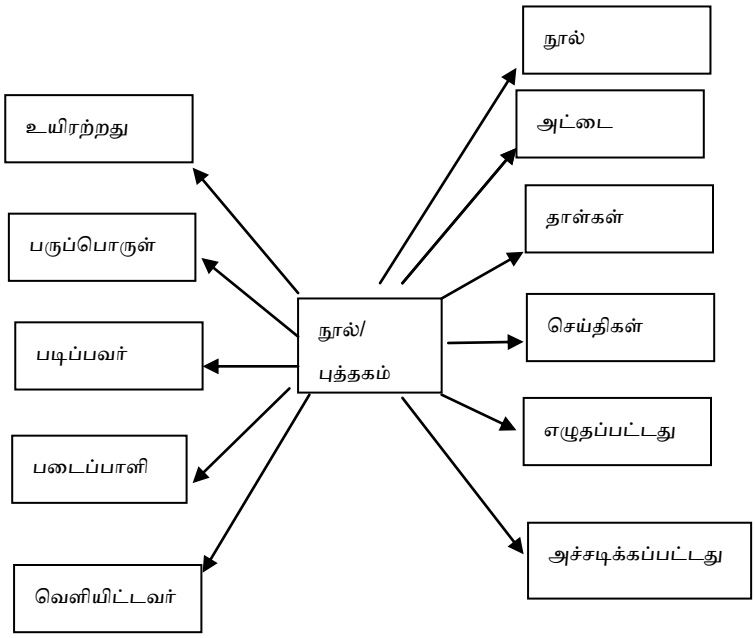
எண்	வாக்கியம்	கருத்துரு
1	தெரசாவின் அறப்பணிகளுக்கு நிகரில்லாத புகழ் மாலைகளாக அமைந்து மகிழ்வித்தன.	மாலை2
2	…தினந்தோறும் காலை மாலை ஆகிய இரு வேளைகளிலும்	மாலை1
3	காலையில் சூரியன் தோன்றிய நேரத்தில் இருந்து மாலையில் மறையும் நேரம் வரை	மாலை1
4	சூரியகாந்திப்பூ காலை, மாலை வேளைகளில் திசை	மாலை1
5	சுற்றுலாப் பேருந்து அடுத்த வெள்ளிக்கிழமை மாலை சென்னைக்குத் திரும்புகிறது	மாலை1
6	காலையில் 5 மணிக்கும், பிறகு 9:30 மணிக்கும், மலையில் 6:30 மணிக்கும் S1	மாலை1
7	அரசு தூக்கிலிட முடிவு செய்த நாளின் முதல் நாள் மலையே, தூக்கிலிட்டுக் S1	மாலை1
8	தத்தும் வீசிய துண்டறிக்கைகள் அன்று மாலையே “இந்துஸ்தான் டைம்ஸ்” S1	மாலை1
9	நடவடிக்கைகளின் காரணமாக நேற்று மாலையிலேயே வந்த செய்தியின்படி S1	மாலை1
10	“குரங்கு கையிலே பூமாலை” என்று அதைப்போல ஊராட்சி S2	மாலை2
11	சிலர், ‘நான் மாலையையே மலர் ஆக்குவேன்	மாலை2
12	அழகு, இளமை, கருவிழி, செல்லம், வாகனம், மாலை, கொடி முதலியவை	(மாலை2)
13	அல்லது திருநீறு காணப்படும் சபையோர் போடும் மாலைகளைப் கழற்றாமலே	மாலை2
14	நானும் அந்தச் செய்தியை நேற்று மாலை அமைச்சரவைக் கூட்டம்	மாலை1
15	உணவுப் பொருள்களை உட்கொள்ளும் பொழுதும், மாலையில் மேய்ச்சலுக்குப் பிறகும்	மாலை1

முந்தைய அணுகுமுறையில் கூறப்பட்ட பொருண்மை மயக்கம் உள்ள *நூல்1* (இழை), *நூல்2* (புத்தகம்) என்ற கருத்துருக்களைக் கருத்துரு வரைபட அகராதி அடிப்படையில் பொருண்மை மயக்கம் நீக்குவதைக் கருதுவோம். *நூல்2இன்* கருத்துரு வரைபட அகராதி முன்னர் தரப்பட்டுள்ளது; இது மீண்டும் இங்கு ஒப்புமைக்காகக் காட்டப்பட்படுள்ளது. *நூல்1*-இன் கருத்துரு வரைபட அகராதி பின்வருமாறு அமையும்.

நூல்1



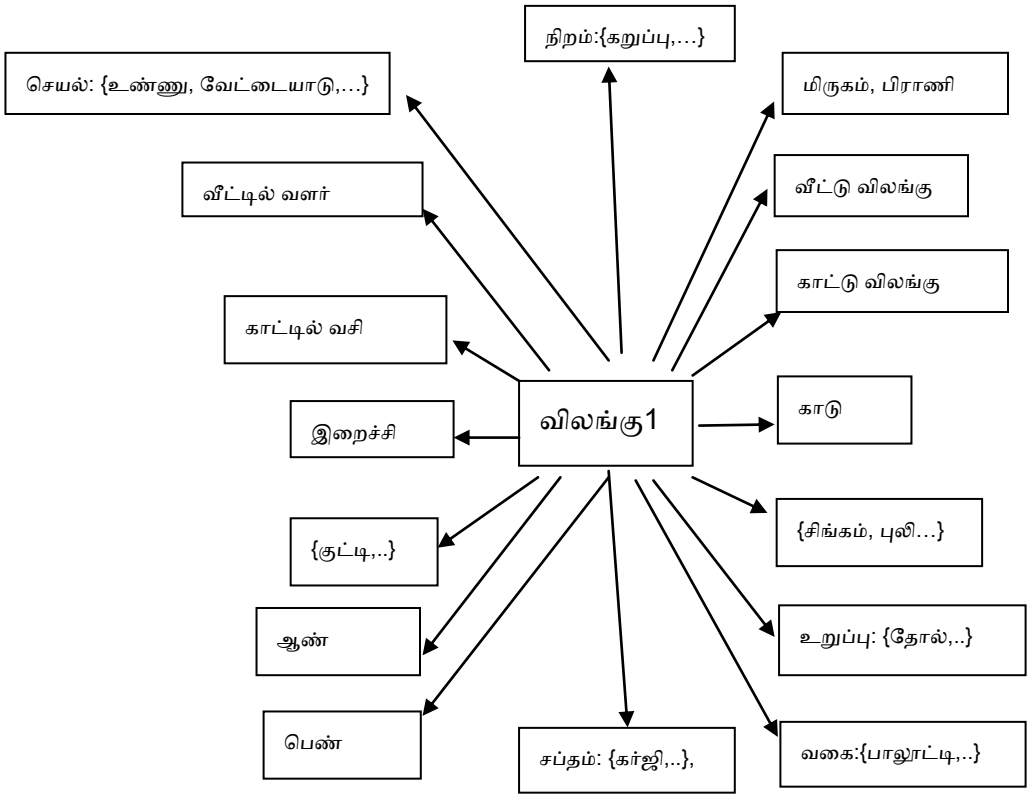
நூல்2



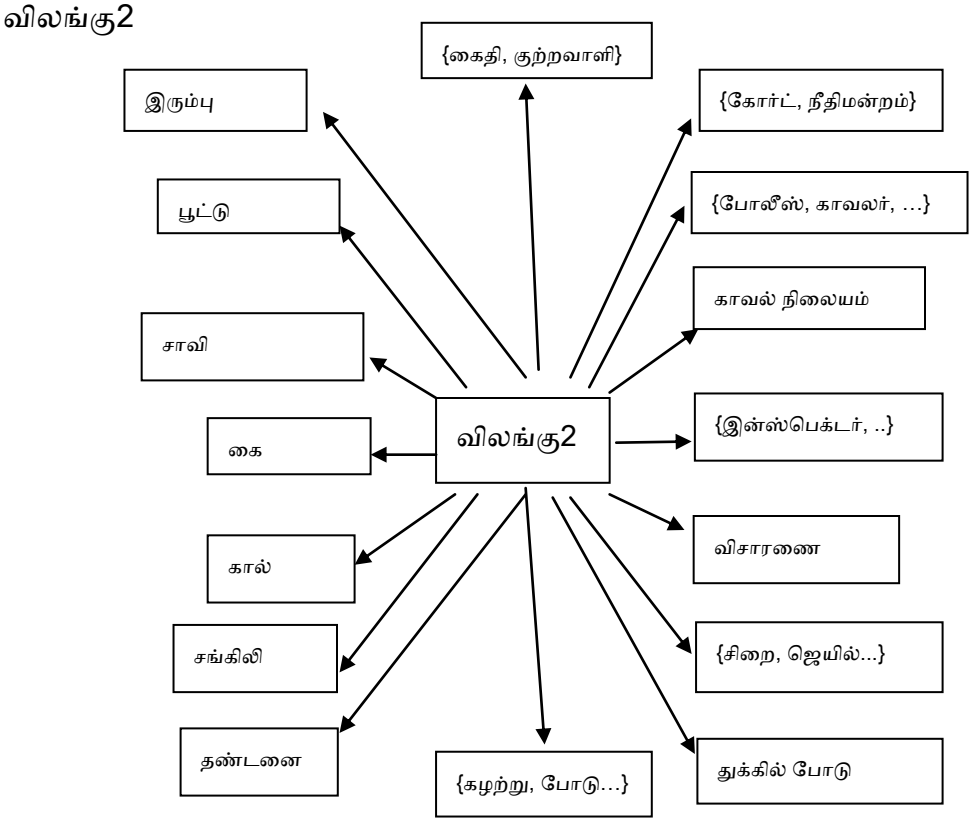
எண்	வாக்கியம்	கருத்துரு
1	… சட்டையின் நூல் இழை….	நூல்1
2	தறியில் நூல் சிக்கி துணி நெய்ய..	நூல்1
3	…நூலில் சாயம் ஏற்றி…	நூல்1
4	கதர் துணி நெய்ய நூல் தேவை…	நூல்1
5	தறியில் ஊடு…நூல்	நூல்1
6	ஊசியில் நூல் கோர்க்க இயல…	நூல்1
7	நூல் ஆசிரியர் பெயர் தெரிவில்லை	நூல்2
8	பழனியப்பா நூல் பதிப்பகத்தில் அச்சடிக்கபட்ட…	நூல்2
9	நூல் 1000 பக்கங்களைக் கொண்டு ….	நூல்2
10	நூல் 1/8 டெம்மி அளவு தாள் கொண்டு …	நூல்2

அடுத்ததாகப் முந்தைய அணுகுமுறையில் பொருண்மை மயக்கம் உள்ளவை என்று எடுத்துக்கொண்ட *விலங்கு* (மிருகம்), *விலங்கு* (கைவிலங்கு) என்ற கருத்துருக்களைக் கருத்துரு வரைபட அகராதி அடிப்படையில் பொருண்மை மயக்கம் நீக்குவதைக் கருதுவோம். *விலங்கு1*, *விலங்கு2* இவற்றின் கருத்துரு வரைபட அகராதி பின்வருமாறு அமையும்

விலங்கு1


நிறம்:{கறுப்பு,…}
செயல்: {உண்ணு, வேட்டையாடு,…}
மிருகம், பிராணி
வீட்டில் வளர்
வீட்டு விலங்கு
காட்டு விலங்கு
காட்டில் வசி
விலங்கு1
இறைச்சி
காடு
{குட்டி,..}
{சிங்கம், புலி…}
ஆண்
உறுப்பு: {தோல்,..}
பெண்
சப்தம்: {கர்ஜி,..},
வகை:{பாலூட்டி,..}

விலங்கு2


{கைதி, குற்றவாளி}
இரும்பு
{கோர்ட், நீதிமன்றம்}
பூட்டு
{போலீஸ், காவலர், …}
காவல் நிலையம்
சாவி
விலங்கு2
கை
{இன்ஸ்பெக்டர், ..}
விசாரணை
கால்
சங்கிலி
{சிறை, ஜெயில்...}
தண்டனை
{கழற்று, போடு…}
துக்கில் போடு

எண்	தரவு வாக்கியங்கள்	கருத்துரு
1	விலங்குகள் காட்டில் அலைந்து திரிந்து…	விலங்கு1
2	காட்டு விலங்கான சிங்கம்…	விலங்கு1
3	காட்டில் சிங்கம், புலி, யானை,…ஆகிய விலங்குகள்..	விலங்கு1
4	விலங்குகளில் பெரிதான யானையின் தந்தங்கள் விலைமிக்கவை என…	விலங்கு1
5	நாய், பூனை, பசு, குதிரை போன்ற வீட்டு விலங்குகளை…	விலங்கு1
6	போலீஸ் குற்றவாளிகளைக் கைது செய்ய…	விலங்கு2
7	கைதியை கோர்ட்டில் கொண்டுபோக…	விலங்கு2
8	கைதிகளை விசாரணை செய்து தண்டனை அளிக்க…	விலங்கு2

உள்ளடங்கு-உள்ளடக்கு மொழி உறவுகள், எதிர்மொழிகள் உறவுகள், முழு-சினை மொழி உறவுகள், வகை உறவுகள், உறுப்பினர் உறவுகள், சொந்த பந்த உறவுகள், சமூக உறவுகள் போன்ற எல்லா உறவுகளையும் கருத்துரு வரைபட அகராதியாகத் தர இயலும்.

5. செம்மொழிப் பனுவல்களில் சொற்பொருண்மை மயக்கநீக்கம்

மேற்சொன்ன இரு அணுகுமுறைகளையும் செம்மொழிப் பனுவல்களில் இடம்பெறும் சொற்களின் சொற்பொருண்மை மயக்கங்களை நீக்கப் பயன்படுத்த இயலும். பின்வருவன இதற்கு அறிவு மூலவளங்களாக அமையும்.

1.சங்க இலக்கியப் பொருட் களஞ்சியம். தமிழ்ப் பல்கலைக் கழகம் தஞ்சாவூர்.

2. பெருஞ்சொல் அகராதி. தமிழ்ப் பல்கலைக் கழகம் தஞ்சாவூர்.

4. சங்க இலக்கியச் சொல்லடைவு
5. சங்க இலக்கிய சொல்லாட்சி அகராதி மற்றும் பொருளடைவு

எடுத்துக்காட்டாகப் புறநானூற்றிலிருந்து எடுக்கப்பட்ட *அணி* என்ற சொல்லின் சொற்பொருளடைவின் மாதிரி பின்வருமாறு பத்து பொருண்மைகளை வேறுபடுத்தி அமையும்.

அணி

1. அழகு

அழகு தாழ்ந்த அணி (புறம் 57.1)

அணிகொள் தேர் (புறம் 68.14)

….

2. அணிகலன்

இழையணிபொலிந்த (புற, 81.1)

3. ஒப்பனை

ஆடு கோட்பாட்டுச் சேரலாதன் போந்தைப் பொழியின் கண் அணியல் (படி 51.9)

4. அணிமை
5. வரிசை
6. திரட்சி
7. ஆடை
8. ஒப்புமை
9. சூழ்தல்
10. பக்கரை (சேணம்)

முன்னர் விளக்கிய சேர்ந்துவருகை அணுகுமுறையைப் பயன்படுத்தி *அணி* என்பதன் பத்து பொருண்மைகளையும் வேற்படுத்தி அறியலாம்.

சங்க இலக்கியப் பொருட்களஞ்சியத்தைப் பயன்படுத்தியும் மேற்சொன்ன அணுகுமுறை அடிப்படையில் சொற்களின் பொருண்மை மயக்கங்களை நீக்கலாம். எடுத்துக்காட்டாக, *களவு* என்ற சொல்லை எடுத்துக்கொண்டால், சங்க இலக்கியப் பொருட்களஞ்சியம் அதற்கு மூன்று பொருண்மைகளை வேறுபடுத்தி அவை வரும் இடங்களையும் பட்டியலிட்டுள்ளது.

1. களவொழுக்கம்

தித்தனின் உறையூரைச் சூழ்ந்த புறங்காடு போல நம் களவும் பல இடையூறுகளைக் கொண்டது (அகம்122:21-23)

நள்ளிரவில் தலைவனுடன் புணர்குறி வாய்த்த களவும் எல்லை கடந்து நீண்டுவிட்டது (அகம் 325:3,4)

….

2. திருட்டு

கொலைத்தொழிலை நீக்கியும் களவை நீக்கியும் உழவு வாழ்க்கையை விரும்பிய வேளாண்மை மாந்தர் வாழ்ந்தனர். (பட்.199, 205)

களவாகிய உழவினைச் செய்யும் கடுங்கண் மழவர் தம் தொழிலுக்கு ஒடுங்காடு செல்வர்.(அகம் 91:11,12).

….

3. களாச்செடி

களவு புளிப்புப்போல விளைந்த கள் புளிப்பேறியிருந்தது (புறம் 328:8)

களவும் பழுத்து நைந்து புளிப்பேறின (அகம் 394:1)

முடிவுரை

சொற்பொருண்மை மயக்கநீக்கத்திற்குப் பல வழிமுறைகள் உள்ளன. கண்காணிக்கபட்ட அணுகுமுறைகளைக் காட்டிலும் கண்காணிக்கப்படாத அணுகுமுறைகள் சிக்கனமானது. அறிவு அடிப்படையிலான அணுகுமுறைகள் வரவேற்கப்படவேண்டியவை ஆகும். மொழியியல் அடிப்படையில், குறிப்பாகச் சொற்பொருண்மையியல் அடிப்படையில் பொருண்மைமயக்கச் சிக்கலுக்குத் தீர்வுகாண முயல்வேண்டும். பொருண்மைமயக்கச் சிக்கல் சொற்பொருண்மை மொழியியல் அடிட்படையில் அணுகப்பட்டால் தான் தீர்வுகாண்பதற்கான வழிமுறைகளை நிறைவாக/தீர்மானமாகக் காண இயலும். கொத்து/திரட்டு நெறிமுறை கண்காணிக்கப்படாத அணுகு முறைகளாகும். இம்முறை ஓரளவுக்குப் பலன் விளைவிப்பதாகும். அவர்தம் வேற்றுமைக் குறியீட்டு முறை அவ்வளவு நம்பத்தகுந்தது அல்ல. தற்போது சொல்வலைகளைப் பயன்படுத்தி பொருண்மை மயக்கநீக்கம் செய்யும் முயற்சிகள் நடந்து வருகின்றன. கருத்துரு வரைபடம் அடிப்படையிலான அணுகுமுறை மொழியில் மற்றும் சொற்பொருண்மையியல் கோட்பாடுகளின் அடிப்படையில் அமைகிறது. மேற்சொன்ன கருத்துரு வரைபட அகராதியை உருவாக்குவது சவாலுக்கு உரியதாகும். இம்முறை சிக்கமானதல்ல. இருப்பினும் இந்நெறிமுறை வரவேற்கத்தகுந்தாகும். கண்காணிக்கப்படாத அணுகுமுறைகளை அல்லது புள்ளியியல்சார் அணுகுமுறைகளைப் பின்பற்றும் முன்னர் அறிவு அடிப்படையிலான சொற்பொருண்மையியல் சார் அணுகுமுறைகளுக்கான ஆயத்தங்கள் மேற்கொள்ளப்படவேண்டும். ஒரு கலப்படமான, அதாவது கண்காணிக்கப்பட்ட, கண்காணிக்கப்படாத மற்றும் அறிவு அடிப்படையிலான அணுகுமுறைகளை இணைத்துச் செய்யப்படும் அணுகுமுறை நம்பத்தகுந்ததாயும் தீர்வுகாண்பதாயும் சிறப்பாகவும் அமைய இயலும்.

துணைநூல்கள்

இராசேந்திரன், ச. 1999. பொருட்புல வகைப்பாடும் சொற்களஞ்சியமும் [Classification into semantic domains and Thesaurus]. புலமை தொகுதி 25, எண்2, டிசம்பர், 1999, 47-66.

இராசேந்திரன், ச. 2001 தற்காலத் தமிழ்சொற்களஞ்சியம். தமிப்பல்கலைக் கழகம், தஞ்சாவூர்.

இராசேந்திரன், ச. 2006. தமிழ் மின்சொற்களஞ்சியம். தமிழ்ப்பல்கலைக் கழகம், தஞ்சாவூர்.

சங்க இலக்கியப் பொருட்களஞ்சியம். தமிழ்ப் பல்கலைக்கழகம்,தஞ்சாவூர்.

பெருஞ்சொல்லகராதி. தமிழிப் பல்கலைக்கழகம், தஞ்சாவூர்.

ARUN, M.C. “Evaluation of Word Sense Disambiguation methods applied on Tamil Corpus. PhD thesis.

BASKARAN, S. 2002. Semantic analyser for word sense disambiguation. MS thesis. Madras Institute of Technology, Anna University, Chennai 2002.

BASKARAN SANKARAN, VAIDEHI V, 2002, "Role of Collocations and Case-Markers inWord Sense Disambiguation: A Clustering-Based Approach". In Proceedings of IEEE International Symposium on Natural Language Processing and Knowledge Engineering 2002. Vol. I. pp. 625-630. Hammamet, Tunisia.

BASKARAN S, VAIDEHI V, 2003, "Collocation based Word Sense Disambiguation using Clustering for Tamil". Communicated to International Journal of Dravidian Linguistics, Thiruvananthapuram, India.

IDE, N. AND V´ERONIS, J. 1998. Word sense disambiguation: The state of the art. Computat. Ling. 24, 1, 1–40.

NAVIGLI, R. 2009. Word Sense Disambiguation: A Survey. ACM Computing Surveys, Vol 41, No.2, Article 10, Publication date February 2009.

PUSTEJOVSKY, J. 1995. The Generative Lexicon. MIT Press, Cambridge, MA.

SOWA, J.F. 1984. Conceptual Structures: Information Processing in Mind and Machine. Reading, Massachusetts: Addison-Wesley Publishing Company.

YAROWSKY, D. 1992. Word-sense disambiguation using statistical models of Roget’s categories trained on large corpora. In Proceedings of the 14th International Conference on Computational Linguistics (COLING, Nantes, France). 454–460.

YAROWSKY, D. 1993. One sense per collocation. In Proceedings of the ARPA Workshop on Human Language Technology (Princeton, NJ). 266–271.

YAROWSKY, D. 1994. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (Las Cruces, NM). 88–95.

YAROWSKY, D. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (Cambridge, MA). 189–196.

YAROWSKY, D. 2000. Hierarchical decision lists for word sense disambiguation. Comput. Human. 34, 1-2, 179–186.

YAROWSKY, D. AND FLORIAN, R. 2002. Evaluating sense disambiguation across diverse parameter spaces. J. Nat. Lang. Eng. 9, 4, 293–310.